

Oy 292 ggcactaacggagacatcatccccggctgacgtacatctcgtgacacgcaacaaagcggc 351
 Db 793 gTSSSSDSTTnCCGCCCMCTCCSTYBMBCYTSTCGGSSSSGAGGYTKCCCGGGCGSS 852
 Oy 352 ggcacgcgtacaggaacgtgacatgtgcgcgaaggagacccccggcc 396
 Db 853 TNGMGTTSSACSSSSSSSVSSSSSKSSAASSSSSVSSSGSSGVS 897

| | |
|------------|---|
| RESULT | 4 |
| CNS0091P/c | |
| LOCUS | CNS0091P |
| DEFINITION | Drosophila melanogaster genome survey sequence TERT3 end of BAC # BAR01D16 of Rpcl-98 library from Drosophila melanogaster (fruit fly) genomic survey sequence. |
| ACCESSION | AL053013 |
| VERSION | AL053013.1 |
| KEYWORDS | GSS. |
| SOURCE | Fruit fly. |
| ORGANISM | Drosophila melanogaster |

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 925)
Genoscope,
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ::

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : net@genoscope.cns.fr)
- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.flyfry.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamooser in Pister de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y?; cn bw sp. the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACpac Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .925 |

| | | | | | |
|-----------------------|-------------------|-----------------|-----------|-------------|------------|
| BASE COUNT | 120 a | 61 c | 61 g | 172 t | 511 others |
| ORIGIN | | | | | |
| Query Match | 12.5% | Score 49.6: | DB 12: | Length 925: | |
| Best Local Similarity | 12.8% | Pred. No. 0.79: | | | |
| Matches 43: | Conservative 157: | Mismatches 136: | Indels 0: | Gaps 0 | |

[illegible]

Dc 720 VASMSGSSSSSSASASSSSSSSSASGACGCCCTGMSGCSCTASMSMAAASSSSSSSSS 661

Qy 300 cggagcatcatccggatgacgtatcgtatcgtgatgccaaccaaagtctgggagagacg 359

Dd 660 CSSMSASSSASSSASSSSSSSGSSSSGACCBMSAGCGGSGVSAASGMSSSVSSG 601

Qy 360 tacagggaaactgacattgcccggaggaaccccggc 395

Dd 600 GRSAGGAGGAGGSSSSSSGSSGSSGSSGAGVCSGSS 565

| RESULT | 5 |
|------------|---|
| BH020987 | |
| LOCUS | BH020987 561 bp DNA linear GSS 25-MAY-2001 |
| DEFINITION | P864c.4LEISHPAC-left 1 Leishmania major Friedlin Cosmid Genomic |
| ACCESSION | BH020987 |
| VERSION | BH020987.1 GI:14202102 |
| KEYWORDS | |
| SOURCE | Leishmania major. |

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--------------------|--|---|--------------------|---------|
| 1 (bases 1 to 561) | Myer, P.J., Vogt, C., Cawthra, J., Klackung, M., Marty, A., Mack, J., Munder, H., Neuyen, D., Robertson, L., Sisk, E., Fazelinia, G., Aggarwal, G., Nelson, S., Seyler, A., Worthey, E. and Stuart, K. | Leishmania major Friedlind Cosmid End Sequences | unpublished (2000) | |
| | Other GSSs: P864c.d, LEISHPAC-right.1 | | | |

TITLE Leishmania major Friedlin Cosmid End Sequences
JOURNAL Unpublished (2000)
COMMENT Other_GSSs: P864c.d_LEISHPAC-right.1
Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@sbri.org
Seq primer: LEISHPAC-left
Class: PAC end.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .561 |

/organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:3664"
 /clone="p864c"
 /clone_11b="Leishmania major Friedlin Cosmid Genomic Library"
 /lab_host="E. coli ED8767"
 /note="Vector: cLHG; Site_1: BamHI; Genomic DNA from Leishmania major Friedlin was partially digested with SauAI, size selected, and ligated with BamHI-digested cLHG cosmid vector DNA. 9216 clones were picked and arrayed. Library construction is described in Ivens et al., Genomics Research, 8:135-145 (1998). The cLHG vector (Acc. No. CV059231) is described in Ryan et al., Gene, 131:145-150 (1993)."

| | | | | |
|-----------------------|-------|---|-------|---------------|
| Query Match | 12.4% | Score 49.2 | DB 12 | Length 561 |
| Best Local Similarity | 63.6% | Pred. No. 0.9 | | |
| Matches | 75 | Conservative | 0 | Mismatches 43 |
| | | | | Indels 0 |
| | | | | Gaps 0 |
| OY | 159 | caagaacgcaacgcgcgcacgaglcacacgctgctcgagagcgtccggcgcaafctc | 218 | |
| | | | | |
| Db | 13 | CCAGATTCGGATCGCGCGCGCTCGTATCCAGGCGCGGTGGGCTCCGCGCTCGGCT | 72 | |
| OY | 219 | cggcatctccacacggcgacgtgatcaacgcggtcgacgcgctccgatcaactcgcc | 276 | |
| | | | | |
| Db | 73 | CGCGATTCACCCGGCGGCGAGCTGATCTCGCCCTTCAACGGCAGCGCGATCAATCAATC | 130 | |
| RESULT | 6 | | | |

[illegible][illegible]

| FEATURES | LOCATION/Qualifiers |
|------------|--|
| RESULTS | 289 bp DNA linear GSS 25-JUL-2000 |
| PE864L | |
| LOCUS | |
| DEFINITION | Leishmania major Friedlin PAC Pe864 left end-sequence, genomic |
| ACCESSION | AL390548 |
| VERSION | AL390548.1 GI:9501524 |
| KEYWORDS | GSS. |
| SOURCE | Leishmania major. |
| ORGANISM | Leishmania major Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; |
| REFERENCE | |
| AUTHORS | Leishmania 1 (bases 1 to 289) Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and Smith, D.F. |
| TITLE | A physical map of the Leishmania major Friedlin genome |
| JOURNAL | Genome Res. 8 (2), 135-145 (1998) |
| MEDLINE | 98146435 |
| REFERENCE | 2 (bases 1 to 289) |
| AUTHORS | Taylor, R.G., Huckle, E.E.J., Ivens, A.C., Rajandream, M.A. and Barrett, B.G. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barretts@sanger.ac.uk and allicats@sanger.ac.uk |
| COMMENT | see http://www.ebi.ac.uk/patases/leish.html Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/Lmajor/ The primer sequence can be obtained from allicats@sanger.ac.uk. |

```

SOURCE
I., 209
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="Eion:5664"
/clone="PAC P864"
BASE COUNT      49 a      86 c     108 g      46 t
ORIGIN

Query Match          12.0% ; Score 47.6; DB 12; Length 289;
Best Local Similarity 62.7%; Pred.No.1.8;
Matches    74; Conservative    0; Mismatches   44; Indels    0; Gaps    0;

Qy    159 caataagcgaaaggcgcacgctcaacgggttgtagggcgtccggcggaagtct 218
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    24 CGAAGCGGCATCGCCGCGGCCCTGTATTATCGAAGGCCCGCGGGTCCCGGGTCGCT 83
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy    219 cgagatctccaccgcgcagcgtgatcacgcgcgctcgcgcgcgtctcgactcgcgc 276
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    84 CCGCATTCACACC CGGGGACGTGATCCTGGCCTTCACAGCGCACGCCGATCAAGTCAGTC 141
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT      8
CNS006XK    935 bp      DNA      linear      GSS 03-JUN-1999
LOCUS
DEFINITION BACR341N09 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066051.1 GI:4945019
VERSION AL066051.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pelegricola; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydriidea; Drosophilidae; Drosophila.
           1 (bases 1 to 935)
Genoscope. Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

```


QY 245 ccgcggtcgcagcgcgtccatcactcgcgcaccgcgatggcgagcgcttaacggc 304
 DB 705 CSAVSSGAASSGSSASGCGSSSGCSCSSSSSMSCSSSSSCGSSCVCSSCVCSS 764
 QY 305 atcatcccgatgacgtatcctcgttgacactgcaacaacgctcgggcgacgctacag 364
 DB 765 CASCBSBSCSCGCASSNASGSSSSSSCSCSVSCMCMCMASASASSSSSSGGSSG 824
 QY 365 ggaacgtgaca 375
 DB 825 SVASGCAV 835

RESULT 12
 BH190076/c 504 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXOC60TF ATXO Arabidopsis thaliana genomic clone ATXOC60, DNA
 DEFINITION sequence.
 ACCESSION BH190076 GI:16302878
 VERSION BH190076
 KEYWORDS GSS.
 SOURCE Chale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 504)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Uutterback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.,
 Survey sequencing of Arabidopsis thaliana BAC T24G4
 JOURNAL Unpublished (2001)
 COMMENT Other-GSS: ATXOC60TR
 CONTACT: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seq primer: TF
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..504
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOC60"
 /clone_1lb="ATXO"
 /note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
 DNA inserted into PHOS2 using BstXI linkers"
 BASE COUNT 73 a 106 c 213 g 110 t 2 others
 ORIGIN

Query Match 11.2%; Score 44.4; DB 12; Length 504;
 Best Local Similarity 53.4%; Pred. No. 8.7;
 Matches 93; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 155 tcgaacaacaagcgcaagcgacagtcgaagcgtgctcgagcgagcgccgcggcgca 214
 DB 401 TCGGCTTCGACGTCATGATGACCAACTCTCCCGGAAGCAAGCTGCGCCGCAAGCCG 342
 QY 215 gtctggcatctccacgcgcgacgtgataccgcggtcgacgcgcctcgatcaactcgg 274
 DB 341 AGATGGCCCTCGCACCTCGCATGATCAACGACTGACTGCTCGAAAAACCGCAAG 282
 QY 275 ccacgcgagtgaggcgagcgcttaacgggcatcatcccggtgacgtcatctcgg 328
 DB 281 CCCCCGTACCGCGGAAGCGGTATCGCCACCTCCACGCAATGTCAGCATGG 228

RESULT 13
 BH189840/c 590 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXOC41TR ATXO Arabidopsis thaliana genomic clone ATXOC41, DNA
 DEFINITION sequence.
 ACCESSION BH189840 GI:16302367
 VERSION BH189840
 KEYWORDS GSS.
 SOURCE Chale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 590)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Uutterback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.,
 Survey sequencing of Arabidopsis thaliana BAC T24G4
 JOURNAL Unpublished (2001)
 COMMENT Other-GSS: ATXOC41TR
 CONTACT: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seq primer: TF
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..590
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOC41"
 /clone_1lb="ATXO"
 /note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
 DNA inserted into PHOS2 using BstXI linkers"
 BASE COUNT 86 a 132 c 251 g 121 t

Query Match 11.2%; Score 44.4; DB 12; Length 590;
 Best Local Similarity 53.4%; Pred. No. 8.8;
 Matches 93; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 155 tcgaacaacaagcgcaagcgacagtcgaagcgtgctcgagcgagcgccgcggcgca 214
 DB 216 TCGGCTTCGACGTCATGATGACCAACTCTCCCGGAAGCAAGCTGCGCCGCAAGCCG 157
 QY 215 gtctggcatctccacgcgcgacgtgataccgcggtcgacgcgcctcgatcaactcgg 274
 DB 156 AGATGGCCCTCGCACCTCGCATGATCAACGACTGACTGCTCGAAAAACCGCAAG 97
 QY 275 ccacgcgagtgaggcgagcgcttaacgggcatcatcccggtgacgtcatctcgg 328
 DB 96 CCCCCGTACCGCGGAAGCGGTATCGCCACCTCCACGCAATGTCAGCATGG 43

RESULT 14
 BH189865 632 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXOD06TF ATXO Arabidopsis thaliana genomic clone ATXOD06, DNA
 DEFINITION sequence.
 ACCESSION BH189865 GI:16302418
 VERSION BH189865
 KEYWORDS GSS.
 SOURCE Chale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 632)

Tue Aug 6 11:59:24 2002

us-09-684-215a-3.rst

Page 9

QY 301 gggcaccatccggtgacgtatccgtgacgtggaacacgaatcgagcgacgct 360
|||||
Db 152284 gggcaccatccggtgacgtatccgtgacgtggaacacgaatcgagcgacgct 152343
QY 361 acaggaacgtgacatggtgacgaggaaccccgcc 396
|||||
Db 152344 acaggaacgtgacatggtgacgaggaaccccgcc 152379

RESULT 2

US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 396; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 3.1e-91;
Matches 396; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 acggcgctgctcgaatacttcacagctgtcccaagggtgaggaagatcgcatccgac 60
|||||
Db 151815 accgcgctgctcgaatacttcacagctgtcccaagggtgaggaagatcgcatccgac 151874
QY 61 gggcagcgatgagcgatcgagcgacgacatccgagtggtgggtgacccacgcttcat 120
|||||
Db 151875 gggcagcgatgagcgatcgagcgacgacatccgagtggtgggtgacccacgcttcat 151934
QY 121 atcgggcctacggccttcctcgtggtgtgtgttcgacacaacagcgcaacgagcaga 180
|||||
Db 151935 atcgggcctacggccttcctcgtggtgtgtgttcgacacaacagcgcaacgagcaga 151994
QY 181 gtccacacggtgtgctggagcgctccggcgcaagtcctcgcatctccacggcgagctg 240
|||||
Db 151995 gtccacacggtgtgctggagcgctccggcgcaagtcctcgcatctccacggcgagctg 152054
QY 241 atcacggcggtcagcgagcgctccgacatccgacacgacgagtgagcgagcgcttaac 300
|||||
Db 152055 atcacggcggtcagcgagcgctccgacatccgacacgacgagtgagcgagcgcttaac 152114
QY 301 gggcaccatccggtgacgtatccgtgacgtggaacacgaatcgagcgacgct 360
|||||
Db 152115 gggcaccatccggtgacgtatccgtgacgtggaacacgaatcgagcgacgct 152174
QY 361 acaggaacgtgacatggtgacgaggaaccccgcc 396
|||||
Db 152175 acaggaacgtgacatggtgacgaggaaccccgcc 152210

RESULT 3
US-08-818-112-4
Sequence 4, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-4

Query Match 99.6%; Score 394.4; DB 4; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.5e-91;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgctgctcgaatacttcacagctgtcccaagggtgaggaagatcgcatccgac 60
|||||
Db 11 acggcgctgctcgaatacttcacagctgtcccaagggtgaggaagatcgcatccgac 70
QY 61 gggcagcgatgagcgatcgagcgacgacatccgagtggtgggtgacccacgcttcat 120
|||||
Db 71 gggcagcgatgagcgatcgagcgacgacatccgagtggtgggtgacccacgcttcat 130
QY 121 atcgggcctacggccttcctcgtggtgtgtgttcgacacaacagcgcaacgagcaga 180
|||||
Db 121 atcgggcctacggccttcctcgtggtgtgtgttcgacacaacagcgcaacgagcaga 190
QY 131 atcgggcctacggccttcctcgtggtgtgtgttcgacacaacagcgcaacgagcaga 190
|||||
Db 131 atcgggcctacggccttcctcgtggtgtgtgttcgacacaacagcgcaacgagcaga 240
QY 181 gtccacacggtgtgctggagcgctccggcgcaagtcctcgcatctccacggcgagctg 240
|||||
Db 191 gtccacacggtgtgctggagcgctccggcgcaagtcctcgcatctccacggcgagctg 250
QY 241 atcacggcggtcagcgagcgctccgacatccgacacgacgagtgagcgagcgcttaac 300
|||||
Db 251 atcacggcggtcagcgagcgctccgacatccgacacgacgagtgagcgagcgcttaac 310
QY 301 gggcaccatccggtgacgtatccgtgacgtggaacacgaatcgagcgacgct 360
|||||
Db 311 gggcaccatccggtgacgtatccgtgacgtggaacacgaatcgagcgacgct 370
QY 361 acaggaacgtgacatggtgacgaggaaccccgcc 396
|||||
Db 371 acaggaacgtgacatggtgacgaggaaccccgcc 406

```
RESULT 4
US-08-818-111-4
; Sequence 4, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Nevo, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-818-111-4

Query Match          99.6%; Score 394.4; DB 4; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.5e-91;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcgctccgataactcagctgtcccaagggtggcaggatcgcacatccgac 60
    |||
DB 11 ACGGCGCGCGTCCGATAACTCAGCTGTCCCAAGGTTGGCGAGGATTCGCGATTCG 70
    |||
QY 61 gggcaggcagatgagctgcggcgagcagatccgagtggtgggggttcaaccacggttcat 120
    |||
DB 71 GGGCAGGCGATGAGCTGCGGCGAGCAGATCCGAGTGGGGGTTCACCCACCGTTTCA 130
    |||
QY 121 atcgggcctacgcgctcctccgcttggtgtgtcgcacacaaggcgaagcgcgaca 180
    |||
DB 131 ATCGGGCCTACGCCCTCCTCCGCTTGgtgtgtcgcacacaaggcgaagcgcgaca 190
    |||
QY 181 gtccaaacgctgtgtcggagagcgtccggcgagcgaagtcgtgcatctccacggcgagctg 240
    |||
DB 191 GTCCAAACGCTGTGTGGAGAGCGCTCCGGCGGCAAGTCTCGCATCTCCACCGGCGAGCTG 250
    |||
QY 241 atcacccggttcgacggcgctcccgatcaactcggccacgcgagtcgagcgagcgttaac 300
    |||
DB 251 ATCACCGGTCGACGGCGCTCCGATCAACTCGGCGCACCGGAGTGGGGAAGCGCTTAAC 310
    |||
QY 301 gggcatcatcccggtgacgtcatctcggtgacgtgcaaaccaagtcggcgagcgagct 360
    |||
```

```
DB 311 GGGCATCATCCCGGTGACGTCATCTCGTGAATGCAACCAACTCGGGGCGACGGCT 370
QY 361 acaggaaactgacatgagccgagggagcccccggcc 396
    |||
DB 371 ACAGGGAACGTGACATTGGCCGAGGAGACCCCGGCGC 406

RESULT 5
US-09-056-556-4
; Sequence 4, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; TREATMENT OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-056-556-4

Query Match          99.6%; Score 394.4; DB 4; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.5e-91;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcgctccgataactcagctgtcccaagggtggcaggatcgcacatccgac 60
    |||
DB 11 ACGGCGCGCGTCCGATAACTCAGCTGTCCCAAGGTTGGCGAGGATTCGCGATTCG 70
    |||
QY 61 gggcaggcagatgagctgcggcgagcagatccgagtggtgggggttcaaccacggttcat 120
    |||
DB 71 GGGCAGGCGATGAGCTGCGGCGAGCAGATCCGAGTGGGGGTTCACCCACCGTTTCA 130
    |||
QY 121 atcgggcctacgcgctcctccgcttggtgtgtcgcacacaaggcgaagcgcgaca 180
    |||
DB 131 ATCGGGCCTACGCCCTCCTCCGCTTGgtgtgtcgcacacaaggcgaagcgcgaca 190
    |||
QY 181 gtccaaacgctgtgtcggagagcgtccggcgagcgaagtcgtgcatctccacggcgagctg 240
    |||
DB 191 GTCCAAACGCTGTGTGGAGAGCGCTCCGGCGGCAAGTCTCGCATCTCCACCGGCGAGCTG 250
    |||
QY 241 atcacccggttcgacggcgctcccgatcaactcggccacgcgagtcgagcgagcgttaac 300
    |||
DB 251 ATCACCGGTCGACGGCGCTCCGATCAACTCGGCGCACCGGAGTGGGGAAGCGCTTAAC 310
    |||
```

| QY | 301 | gggacatccccggatgacgtatctcgtgacgtacgtgacacacaaagtcggagcgacgct | 360 |
|--|-----|--|-----|
| Db | 311 | GGGCACTCATCCCCGGGTACGTCACTCTCGGTAAGTGGCAACCAAGTCGGGCGACCGCT | 370 |
| QY | 361 | acaggaacgtgacattggcgcagagacccccgcc | 396 |
| Db | 371 | ACAGGAAACGTGACATTGGCCGAGGAGACCCCGGCC | 406 |
| RESULT 6 | | | |
| US-08-818-112-17 | | | |
| Sequence 17, Application US/08818112 | | | |
| Patent No. 6290969 | | | |
| GENERAL INFORMATION: | | | |
| APPLICANT: Reed, Steven G. | | | |
| APPLICANT: Skeiky, Yasir A.W. | | | |
| APPLICANT: Dillon, Davin C. | | | |
| APPLICANT: Campos-Neto, Antonio | | | |
| APPLICANT: Houghton, Raymond | | | |
| APPLICANT: Wedvick, Thomas S. | | | |
| APPLICANT: Twardzik, Daniel R. | | | |
| TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY | | | |
| TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS | | | |
| NUMBER OF SEQUENCES: 153 | | | |
| CORRESPONDENCE ADDRESS: | | | |
| ADDRESSEE: SEED AND BERRY LLP | | | |
| STREET: 6300 Columbia Center, 701 Fifth Avenue | | | |
| CITY: Seattle | | | |
| STATE: Washington | | | |
| COUNTRY: USA | | | |
| ZIP: 98104-7092 | | | |
| COMPUTER READABLE FORM: | | | |
| MEDIUM TYPE: Floppy disk | | | |
| COMPUTER: IBM PC compatible | | | |
| OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| SOFTWARE: Patent Release #1.0, Version #1.30 | | | |
| CURRENT APPLICATION DATA: | | | |
| APPLICATION NUMBER: US/08/818,112 | | | |
| FILING DATE: 13-MAR-1997 | | | |
| CLASSIFICATION: 424 | | | |
| ATTORNEY/AGENT INFORMATION: | | | |
| NAME: Makl, David J. | | | |
| REGISTRATION NUMBER: 31,392 | | | |
| REFERENCE/DOCKET NUMBER: 210121.411C6 | | | |
| TELECOMMUNICATION INFORMATION: | | | |
| TELEPHONE: (206) 622-4900 | | | |
| TELEFAX: (206) 682-6031 | | | |
| INFORMATION FOR SEQ ID NO: 17: | | | |
| SEQUENCE CHARACTERISTICS: | | | |
| LENGTH: 1872 base pairs | | | |
| TYPE: nucleic acid | | | |
| STRANDEDNESS: single | | | |
| TOPOLOGY: linear | | | |
| US-08-818-112-17 | | | |
| Query Match 98.8%; Score 391.2; DB 4; Length 1872; | | | |
| Best local Similarity 99.2%; Pred. No. 1.3e-90; | | | |
| Matches 393; Conservative 0; Mismatches 3; Indels 0; Gaps 0; | | | |
| QY | 1 | acggcgcgcgtccgaataactccagctgtccacaggtgtggcaggaattccatcccatc 60 | |
| Db | 758 | ACGGCGCGGCTCGCATATCACTCCAGCGTGTCCAGGGGTGGGCAAGGATTCGCATTCG 817 | |
| QY | 61 | gggcagggcgatggcgatcgcgagcgacatcgatcggtgggggttaaccacggtcat 120 | |
| Db | 818 | GGGCGAGGGCGATGGCATGCGGGGCCAAATCCGATCGGGTGGGGGCGACCCACCGTTCA 877 | |
| QY | 121 | atcgagggcctacgcgcttcctcgagcttgggtgtgttcgaacaacaaggcagagcgaca 180 | |
| Db | 878 | ATCGGGCGCTACCGGCTTCTCGCTGGCTTGGGTGTTCGCAACAACGCGACGCGCAGCA 937 | |
| QY | 181 | gtccacagcgtgtgttcgggaagcgctccgcgcgcaaggtctcgatcctccacggcgacgtg 240 | |

| | | | |
|---|--|---|--|
| Dd | 938 | GTCGCAACCGCTGTGTGGAGAGCGCTCCGGCGGCAGTCTCGGCATCTCCACCGCGGACTG | 997 |
| Oy | 241 | atcacgcggtctcgacggcgcctcgatacaactcggccaccgcgatgtcggaagcgcttaac | 300 |
| Dd | 998 | ATCACCGCGGTGTGGAGAGCGCTCCGATCAACACGCCACCOCGGAGTGCGGACGGCCTTTAAC | 1057 |
| Oy | 301 | gggcatcatcccggttgaaagtcatctcgttgcgttcggccaacaagaatcgggcggaacgct | 360 |
| Dd | 1058 | GGGCATCATCCCGGCTGGAGCTCATCTCGGTGAACHTGGCAAACCAAGTCGGGGGACCGCT | 1117 |
| Oy | 361 | acaagggaacgtgacattgcccggaaggacccccggcc | 396 |
| Dd | 1118 | ACAAGGAACGTGACATTGTGCCGAGGAGACCCCCGGCC | 1153 |
| RESULT 7 US-08-818-111-17 | | | |
| : | Sequence 17, Application US/0818111 | | |
| : | Patent No. 6338852 | | |
| : | GENERAL INFORMATION: | | |
| : | APPLICANT: Reed, Steven G. | | |
| : | APPLICANT: Skeiky, Yasir A.W. | | |
| : | APPLICANT: Dillon, Davin C. | | |
| : | APPLICANT: Campos-Neto, Antonia | | |
| : | APPLICANT: Houghton, Raymond | | |
| : | APPLICANT: Vedvick, Thomas S. | | |
| : | APPLICANT: Twardzik, Daniel R. | | |
| : | TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF | TUBERCULOSIS | |
| : | NUMBER OF SEQUENCES: 148 | | |
| : | CORRESPONDENCE ADDRESS: | | |
| : | ADDRESSEE: SEED and BERRY LLP | | |
| : | STREET: 6300 Columbia Center, 701 Fifth Avenue | | |
| : | City: Seattle | | |
| : | STATE: Washington | | |
| : | COUNTRY: USA | | |
| : | ZIP: 98104-7092 | | |
| : | COMPUTER READABLE FORM: | | |
| : | MEDIUM TYPE: Floppy disk | | |
| : | OPERATING SYSTEM: IBM PC compatible | | |
| : | SOFTWARE: Patentin Release #1.0, Version #1.30 | | |
| : | CURRENT APPLICATION DATA: | | |
| : | APPLICATION NUMBER: US/08/818,111 | | |
| : | FILING DATE: 13-MAR-1997 | | |
| : | CLASSIFICATION: 424 | | |
| : | ATTORNEY/AGENT INFORMATION: | | |
| : | NAME: Makl, David J. | | |
| : | REGISTRATION NUMBER: 31,392 | | |
| : | REFERENCE/DOCKET NUMBER: 210121,417C6 | | |
| : | TELECOMMUNICATION INFORMATION: | | |
| : | TELEPHONE: (206) 622-4900 | | |
| : | TELEFAX: (206) 682-6031 | | |
| : | INFORMATION FOR SEQ ID NO: 17: | | |
| : | SEQUENCE CHARACTERISTICS: | | |
| : | LENGTH: 1872 base pairs | | |
| : | TYPE: nucleic acid | | |
| : | STRANDEDNESS: single | | |
| : | TOPOLOGY: linear | | |
| : | US-08-818-111-17 | | |
| Query Match 98.8%; Score 391.2; DB 4; Length 1872; Best Local Similarity 99.2%; Pred. No. 1,3e-90; | | | |
| Matches | 393; | Conservative 0; | Mismatches 3; Indels 0; Gaps 0; |
| Oy | 1 | acggcgcgtccgatacttcacagctgtcccagggtgggaaggatcgccattccgac | 60 |
| Dd | 758 | ACGGCGCGCTCCGATTAACCTTCCACGCTGTCCCAGCGTGCGCAGGCAATTCGCCATTCGCATC | 817 |
| Oy | 61 | ggggcgagcgatcgagcgatcgagcgagccaatccgatacgggtgtaggggtgataccaccggttat | 120 |
| Dd | 818 | GGGCGAGGCGATCGGCGATCGCGGCCAATTCGATCGGAGTAGGAGGATACCCACCGTTTCAT | 877 |

Matches 177: Conservative 0: Mismatches 175: Indels 3: Gaps 1:

```
QY 34 ggtggcagagatcgcattccgattcggcagcgatgagcgagcagatccga 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1379 GATGCCAGAGCGGCGATCGATCGATCGATCGATCGATCGATCGATCGATCG
QY 94 tcgggtgggggtgaccaccgcttcacatcggtccaccccttcggttggtt 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1439 ATCGCCGAGAGTGTATGATCGACCGGCAAGGCGTCAATCGCTCCCTGGTGTG
QY 154 gtcgacaaacagcgcaac---ggcgacagatccaaacgctgtggtggagcg 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1499 ACCAATGACAAAGACACCCCGGACCAAGATCGTGAAGTGTGCGCGTGTCTGC
QY 211 gcaagctcgcattctcaccgagcgatgataccgagcgagcgagcgagcg 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1559 GCGAAGCGTGAATGCGGAGGCGGCGTGTGACCAAGTGTGACGACGCGGATCAAC
QY 271 tcggcagcgagatgagcgagcgagcgagcgagcgagcgagcgagcgag 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1619 AGCGCGAGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTA
QY 331 acctggcaaacagcgagcgagcgagcgagcgagcgagcgagcgagcgag 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1679 ACCTTTAGAGATCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTA
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 10

US-08-818-111-13

/ Sequence 13, Application US/08818111
/ Patent No. 6338852

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skelky, Yasir A.W.

APPLICANT: Dillon, David C.

APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond

APPLICANT: Weavick, Thomas S.

APPLICANT: Wardik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

NUMBER OF SEQUENCES: 148

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 1771 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-111-13

Query Match 14.9%; Score 59; DB 4; Length 1771;
Best Local Similarity 49.9%; Pred. No. 1.5e-06;

Matches 177: Conservative 0: Mismatches 175: Indels 3: Gaps 1:

```
QY 34 ggtggcagagatcgcattccgattcggcagcgatgagcgagcagatccga 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1379 GATGCCAGAGCGGCGATCGATCGATCGATCGATCGATCGATCGATCGATCG
QY 94 tcgggtgggggtgaccaccgcttcacatcggtccaccccttcggttggtt 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1439 ATCGCCGAGAGTGTATGATCGACCGGCAAGGCGTCAATCGCTCCCTGGTGTG
QY 154 gtcgacaaacagcgcaac---ggcgacagatccaaacgctgtggtggagcg 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1499 ACCAATGACAAAGACACCCCGGACCAAGATCGTGAAGTGTGCGCGTGTCTGC
QY 211 gcaagctcgcattctcaccgagcgatgataccgagcgagcgagcgagcgag 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1559 GCGAAGCGTGAATGCGGAGGCGGCGTGTGACCAAGTGTGACGACGCGGATCAAC
QY 271 tcggcagcgagatgagcgagcgagcgagcgagcgagcgagcgagcgag 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1619 AGCGCGAGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTA
QY 331 acctggcaaacagcgagcgagcgagcgagcgagcgagcgagcgagcgag 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1679 ACCTTTAGAGATCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTA
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11

US-09-056-556-13

/ Sequence 13, Application US/0905556
/ Patent No. 6350456

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skelky, Yasir A.W.

APPLICANT: Dillon, David C.

APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond

APPLICANT: Weavick, Thomas S.

APPLICANT: Wardik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

FILING DATE: 07-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.457

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 1771 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-056-556-13

Query Match 14.9%; Score 59; DB 4; Length 1771;
Best Local Similarity 49.9%; Pred. No. 1.5e-06;

Matches 177: Conservative 0: Mismatches 175: Indels 3: Gaps 1:

Db 55 GACAAAGACACCCGGGCGCCAGATCGTGAAGTAGTGCCGGGTGCTCTCCCGAAC 114
QY 217 CTCGCACTCCCAACGCGACGTGATCAACGCGGCTGCAAGCGGCTCCGATCAACCGCC 276
Db 115 GCTGAGATGCGGAGGCGCGCTGTCACCAAGGTGACGACCGCCGCAACAGCGGG 174
QY 277 ACCGCGATGCGGACGCGCTTAAAGGACATCCCGTGAAGTCACTCCGATGACCGG 336
Db 175 GACGCGTGTGGTCCCGCGCGTCCCAAGCGCGCGCGCAAGGTGGCGTAACTTTT 234
QY 337 CAACAACAGTCGCGGCGGCGGCTACAGGAGGAGCGTACATGTCGCGGAG 385
Db 235 CAGATCCCTCGGGCGGTAGCCGACAGTGAAGTACCCCTCGGCAAG 283

RESULT 14
US-09-056-556-10
Sequence 10, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Marki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-10

Query Match 14.6%; Score 57.8; DB 4; Length 622;
Best Local Similarity 53.3%; Pred. No. 2, 4e-06;
Matches 122; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 157 GACAACAAGGCAAGCAAGTCAAGCGTGTGCGGAGCGCTCCGCGGCAAGT 216
Db 55 GACAAAGACACCCGGGCGCCAGATCGTGAAGTAGTGCCGGGTGCTCTCCCGAAC 114
QY 217 CTCGCACTCCCAACGCGACGTGATCAACGCGGCTGCAAGCGGCTCCGATCAACCGCC 276
Db 115 GCTGAGATGCGGAGGCGCGCTGTCACCAAGGTGACGACCGCCGCAACAGCGGG 174
QY 277 ACCGCGATGCGGACGCGCTTAAAGGACATCCCGTGAAGTCACTCCGATGACCGG 336
Db 175 GACGCGTGTGGTCCCGCGCGTCCCAAGCGCGCGCGCAAGGTGGCGTAACTTTT 234

QY 337 CAACAACAGTCGCGGCGGCGGCTACAGGAGGAGCGTACATGTCGCGGAG 385
Db 235 CAGATCCCTCGGGCGGTAGCCGACAGTGAAGTACCCCTCGGCAAG 283

RESULT 15
US-09-199-637A-131
Sequence 131, Application US/09199637A
Patent No. 635411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wah
APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
FILE REFERENCE: 00786/361002
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 131
LENGTH: 1436
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-131

Query Match 11.4%; Score 45.2; DB 4; Length 1436;
Best Local Similarity 52.7%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 139 CTCGCTGCGGCTGCTGCTGCGAACAAGCGGCAAGCGGCGGCAAGCGGCTGCGG 198
Db 903 CTCGCGAGTCTCGGCTGCGGCTGCGGCAAGCGGCTGCGGCTGCGGCTGCGG 962
QY 199 AGCGCTCGGCGGCAAGTCTCGGCTGCGGCAAGCGGCTGCGGCTGCGGCTGCGG 258
Db 963 GACGCTCGGCGGCAAGTCTCGGCTGCGGCAAGCGGCTGCGGCTGCGGCTGCGG 1022
QY 259 GCTCGATCAACTCGGCGGCAAGTCTCGGCTGCGGCAAGCGGCTGCGGCTGCGG 318
Db 1023 CAGTCAAGCAAGTCTCGGCTGCGGCAAGTCTCGGCTGCGGCTGCGGCTGCGG 1082
QY 319 GTCACT 324
Db 1083 AAGATC 1088

Search completed: August 6, 2002, 11:07:45
Job time: 8085 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 08:56:55 ; Search time 1841.62 Seconds
(without alignments)
4499.793 Million cell updates/sec

US-09-684-215A-3
396
1 acgcgcgcgtccgataact.....tgcgcgagggagaccccgcc 396

Title: US-09-684-215A-3
Sequence: 1 acgcgcgcgtccgataact.....tgcgcgagggagaccccgcc 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match Length | DB ID | Description |
|------------|-------|-------|--------------|-------|-------------|
|------------|-------|-------|--------------|-------|-------------|

| | | | | | | |
|----|-------|-------|--------|----|-----------|---------------------|
| 1 | 396 | 100.0 | 1068 | 6 | AX005788 | AX005788 Sequence |
| 2 | 396 | 100.0 | 1143 | 6 | AX005790 | AX005790 Sequence |
| 3 | 396 | 100.0 | 11700 | 1 | MFC1418B | 296071 Mycobacteri |
| 4 | 396 | 100.0 | 14029 | 1 | AE006925 | AE006925 Mycobacte |
| 5 | 394.4 | 99.6 | 447 | 6 | AR169152 | AR169152 Sequence |
| 6 | 394.4 | 99.6 | 447 | 6 | BD006325 | BD006325 Compounds |
| 7 | 394.4 | 99.6 | 447 | 6 | BD006445 | BD006445 Compounds |
| 8 | 391.2 | 98.8 | 1872 | 6 | AR169165 | AR169165 Sequence |
| 9 | 391.2 | 98.8 | 1872 | 6 | BD006338 | BD006338 Compounds |
| 10 | 391.2 | 98.8 | 1872 | 6 | BD006458 | BD006458 Compounds |
| 11 | 354 | 89.4 | 675 | 6 | AX201049 | AX201049 Sequence |
| 12 | 354 | 89.4 | 675 | 6 | AX267848 | AX267848 Sequence |
| 13 | 354 | 89.4 | 675 | 6 | AX351489 | AX351489 Sequence |
| 14 | 354 | 89.4 | 915 | 6 | AX201061 | AX201061 Sequence |
| 15 | 354 | 89.4 | 915 | 6 | AX267860 | AX267860 Sequence |
| 16 | 354 | 89.4 | 1203 | 6 | AX201078 | AX201078 Sequence |
| 17 | 354 | 89.4 | 1203 | 6 | AX267877 | AX267877 Sequence |
| 18 | 354 | 89.4 | 1464 | 6 | AX156105 | AX156105 Sequence |
| 19 | 354 | 89.4 | 1557 | 6 | AX156089 | AX156089 Sequence |
| 20 | 354 | 89.4 | 1578 | 6 | AX155945 | AX155945 Sequence |
| 21 | 354 | 89.4 | 1590 | 6 | AX156986 | AX156986 Sequence |
| 22 | 354 | 89.4 | 1752 | 6 | AX156109 | AX156109 Sequence |
| 23 | 354 | 89.4 | 1752 | 6 | AX156093 | AX156093 Sequence |
| 24 | 354 | 89.4 | 1860 | 6 | AX156065 | AX156065 Sequence |
| 25 | 354 | 89.4 | 1896 | 6 | AX156081 | AX156081 Sequence |
| 26 | 354 | 89.4 | 1941 | 6 | AX156073 | AX156073 Sequence |
| 27 | 354 | 89.4 | 1965 | 6 | AX156097 | AX156097 Sequence |
| 28 | 354 | 89.4 | 2052 | 6 | AX156113 | AX156113 Sequence |
| 29 | 354 | 89.4 | 2076 | 6 | AX156069 | AX156069 Sequence |
| 30 | 354 | 89.4 | 2103 | 6 | AX156101 | AX156101 Sequence |
| 31 | 354 | 89.4 | 2148 | 6 | AX156077 | AX156077 Sequence |
| 32 | 354 | 89.4 | 2148 | 6 | AX156085 | AX156085 Sequence |
| 33 | 354 | 89.4 | 2148 | 6 | AX156085 | AX156085 Sequence |
| 34 | 199.2 | 50.3 | 1110 | 1 | MPJ4KDPAG | Z23092 M.paratuber |
| 35 | 187 | 47.2 | 37040 | 1 | MSGB26CS | L7816 Mycobacteri |
| 36 | 68.6 | 17.3 | 269203 | 1 | MLEPRN10 | AL583926 Mycobacte |
| 37 | 68.6 | 17.3 | 37304 | 1 | MLECL373 | AL583917 Mycobacte |
| 38 | 62.2 | 15.7 | 134426 | 14 | HLICG | M75136 Ictialurid h |
| 39 | 59 | 14.9 | 1395 | 6 | AX005764 | AX005764 Sequence |
| 40 | 59 | 14.9 | 1771 | 6 | AR169161 | AR169161 Sequence |
| 41 | 59 | 14.9 | 1771 | 6 | BD006334 | BD006334 Compounds |
| 42 | 59 | 14.9 | 1771 | 6 | BD006454 | BD006454 Compounds |
| 43 | 59 | 14.9 | 1812 | 6 | AX005766 | AX005766 Sequence |
| 44 | 59 | 14.9 | 16150 | 1 | MPV044 | AL021999 Mycobacte |
| 45 | 59 | 14.9 | 16168 | 1 | AE006985 | AE006985 Mycobacte |

ALIGNMENTS

| RESULT | 1 | 1068 bp | DNA | linear | PAT 24-AUG-2000 |
|------------|---|---------|-----|--------|-----------------|
| AX005788 | Sequence 907 from Patent WO9909186. | | | | |
| AX005788 | AX005788 | | | | |
| AX005788.1 | GR:9928795 | | | | |
| ORGANISM | Mycobacterium tuberculosis. | | | | |
| SOURCE | Mycobacterium tuberculosis | | | | |
| REFERENCE | Portnoi, D. and Guigueno, A. | | | | |
| AUTHORS | Polypeptide nucleic sequences exported from mycobacteria, vectors | | | | |
| TITLE | comprising same and uses for diagnosing and preventing tuberculosis | | | | |
| JOURNAL | Patent: WO 9909186-A 907 25-FEB-1999; | | | | |
| FEATURES | Portnoi DENIS (FR); GUIGUENO AGNES (FR) | | | | |
| Source | Location/Qualifiers | | | | |
| CDS | 1..1068 | | | | |
| | /organism="Mycobacterium tuberculosis" | | | | |
| | /db_xref="taxon:1773" | | | | |
| | 1..1068 | | | | |

```
/note="50D"  
/codon_start=1  
/transl_table=11  
/protein_id="CAC05178.1"  
/db_xref="GI:9928796"  
/translation="MSNSRRSLRMSWLSVLAAGLGLATAPAOAAPPALSDREAP  
PALPLDPSAMVAOVPQVAVNTKLGNNAGAGTGIYIDNGVVLITNNHVLGATID  
INAFVSGQTVGVVGVVADVDVAVLQIRAGAGLPSAIGGVAVPGVAVMAGNSG  
GGGTPRAVPGVAVVAGTVOASDVLTAETLNLGIOPDAIOPDSDSGPVVNGLSG  
VGGMTAASDNFOLSGGGGFAPIPIGOAMAIAGIOTRSSGGSPVTHIGPTAFLGIVD  
NNGARVORVYVGSAPASLSTGISTGVITAVGAPRINSATMADLNGHHPDVISVT  
WOTKSGCTRTGNTVAEGRPA"
```

BASE COUNT 169 a 352 c 375 g 172 t
ORIGIN

Query Match 100.0%; Score 396; DB 6; Length 1068;
Best Local Similarity 100.0%; Pred. No. 7.4e-56;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 acggccgctcgcataactcagctcgtcccaaggtgagcagagatcgcattccgac 60  
DB 670 ACggccGctcgcAtaactCAgctcgtccCAaggtGagcAgagatcgcAttccGac 729  
QY 62 gggcagcgatgagcagatcgcagcagatcgcagtgagtgagtgacaccacgttcat 120  
DB 730 GggcAGcgATgagCAgAtcgcAGatcgcAGtgAgagtgAcaccACgtTCat 789  
QY 121 atcgagcctacgccttcctcgtcgttggtgtgtcgaacaacagcgagcagca 180  
DB 790 AtcgGcctAcgccttcctcGtgcgttgGtGtgcAaCAacAGcgAGcagCA 849  
QY 181 gtccaaagcgtgagtcgagcagcagcagcagcagcagcagcagcagcagcagc 240  
DB 850 GtccAAagcGtgagtcGagcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgc 909  
QY 241 ataccgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
DB 910 AtAcCGagtgAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgc 969  
QY 301 gggcatalcctccggtgagcagcagcagcagcagcagcagcagcagcagcagc 360  
DB 970 GggcAtalcctccGgtgagCAgCagcAgcAgcAgcAgcAgcAgcAgcAgc 1029  
QY 361 acaggaagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 396  
DB 1030 ACaGgaAGcGtgAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgc 1065
```

RESULT 2
AX005790 1143 bp DNA linear PAT 24-AUG-2000
LOCUS
DEFINITION Sequence 909 from Patent WO9909186.
ACCESSION AX005790
VERSION AX005790.1 GI:9928797
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE
AUTHORS Portnoi, D. and Guigueno, A.
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors
comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL Patent: WO 9909186-A 909 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source
1. .1143
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
CDS
/note="SEQ ID NO 50F"

```
/codon_start=1  
/transl_table=11  
/protein_id="CAC05179.1"  
/db_xref="GI:9928798"  
/translation="KNPAARTLAKACTISDSDYDGTGCGYNSRRSLRMSWLSVLA  
VGLGLATAPAOAAPPALSDREAPDPALEPLDPSAMVAOVPQVAVNTKLGNNVGA  
GTGIVYIDNGVVLITNNHVLGATIDINAFVSGQTVGVVGVVADVDVAVLQIRAG  
GLPSAIGGVAVPGVAVMAGNSGGGTPRAVPGVAVVAGTVOASDVLTAETLNLGI  
OPDAIOPDSDSGPVVNGLSG VGGMTAASDNFOLSGGGGFAPIPIGOAMAIAGI  
OTRSSGGSPVTHIGPTAFLGIVD NNGARVORVYVGSAPASLSTGISTGVITAVG  
APRINSATMADLNGHHPDVISVTWOTKSGCTRTGNTVAEGRPA"
```

BASE COUNT 189 a 373 c 395 g 186 t
ORIGIN

Query Match 100.0%; Score 396; DB 6; Length 1143;
Best Local Similarity 100.0%; Pred. No. 7.4e-56;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 acggccgctcgcataactcagctcgtcccaaggtgagcagagatcgcattccgac 60  
DB 745 ACggccGctcgcAtaactCAgctcgtccCAaggtGagcAgagatcgcAttccGac 804  
QY 61 gggcagcgatgagcagatcgcagcagatcgcagtgagtgagtgacaccacgttcat 120  
DB 805 GggcAGcgATgagCAgAtcgcAGatcgcAGtgAgagtgAcaccACgtTCat 864  
QY 121 atcgagcctacgccttcctcgtcgttggtgtgtcgaacaacagcgagcagca 180  
DB 865 AtcgGcctAcgccttcctcGtgcgttgGtGtgcAaCAacAGcgAGcagCA 924  
QY 181 gtccaaagcgtgagtcgagcagcagcagcagcagcagcagcagcagcagcagc 240  
DB 925 GtccAAagcGtgagtcGagcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgc 984  
QY 241 ataccgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
DB 985 AtAcCGagtgAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgc 1044  
QY 301 gggcatalcctccggtgagcagcagcagcagcagcagcagcagcagcagcagc 360  
DB 1045 GggcAtalcctccGgtgagCAgCagcAgcAgcAgcAgcAgcAgcAgcAgc 1104  
QY 361 acaggaagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 396  
DB 1105 ACaGgaAGcGtgAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgcAgc 1140
```

RESULT 3
MTC1418B 11700 bp DNA linear BCT 03-AUG-2001
LOCUS
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 7/162.
ACCESSION Z96071 AL123456
VERSION Z96071.1 GI:3242254
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltham, D., Gentles, S.,
Hamlin, N., Holt, R., Jones, S., Kaul, R., Kirov, A., Kitchin, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S., and
Barrell, B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature. 393 (6685), 537-544 (1998)

MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL98295987
2 (bases 1 to 11700)
Parkhill, J.

Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB0 1SA unite de Genetique Moleculaire Bacterieme, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 20, 1998 this sequence version replaced gi:2181960.

COMMENT

Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in Tlpase (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

source

1. 11700
/organism="Mycobacterium tuberculosis H37Rv"

/strain="H37Rv"

/db_xref="taxon:83332"

misc_feature

/note="fragment designated v031. Does not represent a physical clone"

RBS

/note="possible RBS, AAGGAG, for Rv0119"

gene

CDS

97. 1674
/gene="fadD7"

/note="Rv0119, (MTV031.13-MTC1418B.01), len: 525.fadD7, possible coenzyme A-ligase similar to 4-comurate:COA ligase of many organisms e.g. PFU39405.1 (537 aa). Contains PS00455 putative AMP-binding domain signature. FASTA scores: gplU39405|PFU39405.1 Plus taeda xylem 4-comurate:COA (537 aa) opt:4832-score: 526.1 E(): 8.3e-22: 28.2% identity in 440 aa overlap score is 0.896"

/codon_start=1

/transl_table=11

/product="fadD7"

/protein_id="CAB09455.1"

/db_xref="GI:3242255"

/translation="MASDFGPRIDLVVNAATRLPEAPALVVTADIAISHRDALRY DELAGLTRSGILPGDRVALRWGSNAEFVVALIASRDLVVPDLADPTTEORVRS QAAGRVYLIDPDGPHDRAEPTTRKMPITLVYVGGSGSGTSLSHDAATERNPATS TPEGRLPDAMTETGGTGLPKVPMWTHANASSVRAITIGYRLSPDAIVAAWPLY HGHGILASILASGAVSLPARFSAHFTWDIKAVGATWTAVPETIHOILLERS ATEPSGRPAALRTIRSCSAPLTAQAALALQEPFAVCAFGMTAEATHOVTTQIEG IDQETPVYSTGLVSTGAQIRIVSGDILPLPAGAVEIMRGTTVARGVLDPTIT ANFTDGMRLTGDLGSLAGDLSIRGIRKELINGGEKISPERVEGVLASHPNMEA AVFGVPHOLGVAVAIVYPRNSAPPTREELVQFORHLAEFAISQFENSGLPHTA KSLDRRAVAERFGHSV"

610. 645
/gene="fadD7"

/note="PS00455 Putative AMP-binding domain signature, [LIVMFY].[2][STG][STAG][GIST][STETI][SC].[PASTLIVM][KR], info count = 22.0"

1309. 11293
/organism="Mycobacterium tuberculosis H37Rv"

/strain="H37Rv"

/db_xref="taxon:83332"

/clone="1418B"

misc_feature

610. 645
/gene="fadD7"

/note="PS00455 Putative AMP-binding domain signature, [LIVMFY].[2][STG][STAG][GIST][STETI][SC].[PASTLIVM][KR], info count = 22.0"

1309. 11293
/organism="Mycobacterium tuberculosis H37Rv"

/strain="H37Rv"

/db_xref="taxon:83332"

/clone="1418B"

/gene="fadD7"

/note="PS00455 Putative AMP-binding domain signature, [LIVMFY].[2][STG][STAG][GIST][STETI][SC].[PASTLIVM][KR], info count = 22.0"

1309. 11293
/organism="Mycobacterium tuberculosis H37Rv"

/strain="H37Rv"

/db_xref="taxon:83332"

/clone="1418B"

misc_feature

610. 645
/gene="fadD7"

/note="PS00455 Putative AMP-binding domain signature, [LIVMFY].[2][STG][STAG][GIST][STETI][SC].[PASTLIVM][KR], info count = 22.0"

1309. 11293
/organism="Mycobacterium tuberculosis H37Rv"

/strain="H37Rv"

/db_xref="taxon:83332"

/clone="1418B"

gene

CDS

complement(1675..3819)
/gene="fusA2"complement(1675..3819)
/gene="fusA2"

/note="Rv0120c, (MTC1418B.02c), fusA2, len: 714. Elongation factor G, similar to eg. EFG ECOLI P02996 elongation factor G (ef-g). Escherichia coli (703 aa), fasta scores: opt:1049 z-score: 1105.5 E(): 0.32.5%

identity in 717 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop). Also similar to M. tuberculosis fusA, MTCY210.01 (39.1% identity in 299 aa overlap). Similar to sp|P30767|EFG_MYCE ELONGATION FACTOR G (EF-G). (701 aa): 31.7% identity in 710 aa overlap."

RBS

gene

CDS

/codon_start=1
/transl_table=11
/product="fusA2"

/protein_id="CAB09448.1"

/db_xref="GI:2181962"

/translation="MADRVNASGAAAPFANGPGVRRVNVLPVSGGKRTLLLEALL VAAKVLSPGSEVTGTCDEDEAETROQSVGLAVALADGIKVNLVDTGADVF GELRAGLRADCAFLVYANEGYDEPTSLMOGSCOMPRAVITIKLIDHARANYREA LTAQADAFGRKVLVLYPSGDLIGLSOLAYEADGRRTTPRPAESDTEIEARGA LIRGTLRESPEDESMERYLGGPGLIDSLVLODEKAVARGSPFPVTPVSGVCTLE LLEVATRGFPSPMEHRLPEVFTTPOGPHAEACDNDARFLAEVYTKTSPYGRVSLY RVESGTRIPDTYHVSCHSFFGGGNTNHPHDEBERGLVSLFPGKQORPAALV AGDICALGKLSRAEVTGDTLSDAEPLVAKWMPDEPLTAIAAARKTDEKSLVGLG RLAAEDPTLRIEONQETHOVLMCGEBAHAGVLDLANKYGVSTVTELRPLRETF AGNAKGGRHRIKOSGGHGYCDIEVEPIEGSGEFELDKVVGCAVAPQPIPNNEKG VRAOMDGVHAGVPVVDIRVTLIDGKHSYDSDFAFOMAGALALREAAATVILILE PIDEISLVDPDDFVGAVLSDLSRRGVLGCTETAGHDRIYKAKEVPQVELTFRALDLR SLAHGASFTRSFARYEMPESSAARVAKAGAC"

complement(3709..3732)
/gene="fusA2"/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(3956..4390)
/gene="Rv0121c"complement(3956..4390)
/gene="Rv0121c"

/note="Rv0121c, (MTC1418B.03c), unknown, len: 144 aa"

misc_feature

gene

CDS

97. 1674
/gene="fadD7"

/note="Rv0119, (MTV031.13-MTC1418B.01), len: 525.fadD7, possible coenzyme A-ligase similar to 4-comurate:COA ligase of many organisms e.g. PFU39405.1 (537 aa). Contains PS00455 putative AMP-binding domain signature. FASTA scores: gplU39405|PFU39405.1 Plus taeda xylem 4-comurate:COA (537 aa) opt:4832-score: 526.1 E(): 8.3e-22: 28.2% identity in 440 aa overlap score is 0.896"

/codon_start=1

/transl_table=11

/product="hypothetical protein Rv0121c"

/protein_id="CAB09449.1"

/db_xref="GI:2181963"

/translation="MGEFDPKRLRPAOSPVARIATSPDGPVPLVVFALGARPAEA TGAIVYTAADAKRKTQRLRLANENHRASVLVDSYADMTQLMWRADSVAAIH RDGEVRRARVRLERAKTAOYOSVPLNPVTAIVORASNAHA"

complement(4399..4401)
/note="possible RBS upstream of Rv0121c"4532..4535
/note="possible RBS upstream of Rv0121c"4539..4907
/note="possible RBS upstream of Rv0122"4539..4907
/note="Rv0122"

/gene="Rv0122"

/note="Rv0122, (MTC1418B.04), unknown, len: 133 aa"

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0122"

/protein_id="CAB09450.1"

/db_xref="GI:2181964"

/translation="MAGSVSAAGIGVNGVNTETNRDQCYRVERPTVDALHPREYR HTKGVQVRYTRNARKHRYVAMRGVAVIOEDSLVYQGRDTSGRLTJEVVA VEADGDLITTHAMPKEMKR"

4890..4895
/gene="Rv0122"

/note="possible RBS upstream of Rv0123"

4904..5272
/gene="Rv0123"

CDS

```

/ gene="RV0123"
/ note="RV0123, (MTC1418B.05), unknown, len: 133 aa"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein RV0123"
/ protein_id="CAB09451.1"
/ db_xref="GI:2181965"
/ db_xref="SPTREMBL:007173"
/ translation="MTKKPRNPADYVIGDDVEVSDVLDKOEYVVDGRLTDEREOM
ASESLRLAREANLIPGKSLSGSAHSPAVYVSKATHAKKELARSKMSVSKL
LRPVLDREVOREGRIILPRR"
RBS
5570..5574
/ note="possible RBS upstream of RV0124"
gene
5581..7044
/ gene="PE_PGRS"
5581..7044
/ gene="PE_PGRS"
5581..7044
/ gene="PE_PGRS"
/ note="RV0124, (MTC1418B.06), len: 487. Member of the
PGRS family of glycine-rich proteins, highly similar to
eg. Y0DP_MYCTU_Q50615 hypothetical glycine-rich 40.8 kd
protein (498 aa), fasta scores; opt: 1730 z-score: 1201.5
E(): 0, 60.7% identity in 504 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="PE_PGRS"
/ protein_id="CAB09452.1"
/ db_xref="GI:2181966"
/ db_xref="SPTREMBL:007174"
/ translation="MSFVSVAPELVVAAATDAGISASIAANAAAAAPTAVIAGA
DEVSATIAALFSGHAYQALSNQAAPFOOTLGGAGATRAAEQVEOCLLAI
NAPYQALGRPLINGADGAGTGAAGAGLILYNGNGSGSAGAGSAGAGLIT
Query Match 100.0%; Score 396; DB 1; Length 11700;
Best Local Similarity 100.0%; Pred. No. 5.1e-56;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.
1 (bases 1 to 14029)
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 14029)
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-Apr-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..14029
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/ db_xref="taxon:83331"
/ note="clinical strain"
complement(99..533)
/ gene="MT0129"
complement(99..533)
/ gene="MT0129"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AKK44353.1"
/ db_xref="GI:13879611"
/ db_xref="GI:13879612"
/ translation="MGEFDPKLRFAQSPVARLSTPDGTPLVPPVAFGARRPAA
TGADVITYAVDARKRTQRLRLANLEHNPASVYVSDADMTQLMWWVRADGVAAIH
RDGEVRAAYRLRLAKRYTOYOSVPLNGPVIAIAYQRMASRHA"
complement(643..1062)
/ gene="MT0130"
complement(643..1062)
/ gene="MT0130"
complement(643..1062)
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AKK44354.1"
/ db_xref="GI:13879612"
/ translation="MASSSSPLRPHCVSDQVYVGPDDIGKTAARTAAIVORA
IFLNDNRNVAHSGDVAIVGAVVDFGVRRDPPPLHASSMYSITLKCOSVNGRPDAIL
LVSVRICHVQYDPTDSCGGRDPQLPCAPLDYHRHH"
1047..1415
/ gene="MT0131"
1047..1415
/ gene="MT0131"
/ note="identified by match to protein family HMM"
/ codon_start=1
/ transl_table=11
/ product="DNA-binding protein, CopG family"
/ protein_id="AKK44355.1"
/ db_xref="GI:13879613"
/ db_xref="GI:13879613"
/ translation="MTKKPRNPADYVIGDDVEVSDVLDKOEYVVDGRLTDEREOM
ASESLRLAREANLIPGKSLSGSAHSPAVYVSKATHAKKELARSKMSVSKL
LRPVLDREVOREGRIILPRR"
1682..3367
/ gene="MT0132"
1682..3367
/ gene="MT0132"
/ note="similar to GB:AL123456; identified by sequence

```



```

QY 301 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 360
    |||||||
Db 4488 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 4547
QY 361 acagggaacgtgacatctgcccgaaggaccccgcc 396
    |||||||
Db 4548 acagggaacgtgacatctgcccgaaggaccccgcc 4583

RESULT 5
AR169152 447 bp DNA linear PAT 17-DEC-2001
LOCUS AR169152
DEFINITION Sequence 4 from patent US 6290969.
ACCESSION AR169152
VERSION AR169152.1 GI:17906927
KEYWORDS
SOURCE Unknown.
ORGANISM Unidentified.
REFERENCE 1 (bases 1 to 447)
AUTHORS Reed,S.G., Skelky,Y.A.W., Dillon,D.C., Campos-Neto,A., Houghton,R.,
          Vedvik,T.S. and Twardzik,D.R.
TITLE Compounds and methods for immunotherapy and diagnosis of
        tuberculosis
JOURNAL Patent: US 6290969-A 4 18-SEP-2001;
FEATURES Location/Qualifiers
source 1..447
BASE COUNT 79 a 146 c 149 g 72 t 1 others
ORIGIN

Query Match 99.6%; Score 394.4; DB 6; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.6e-55;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcgctcgatatacttccagctgctccagggcgagggatcgccattccgac 60
    |||||||
Db 11 acggcgcgctcgatatacttccagctgctccagggcgagggatcgccattccgac 70
QY 61 gggcagcgatgacatcgccgagcagatccgacatcggtggggggtcaccacgcttcat 120
    |||||||
Db 71 gggcagcgatgacatcgccgagcagatccgacatcggtggggggtcaccacgcttcat 130
QY 121 atcgagcctacgcgcttccctcgctggtgtgttcgacaaacaggaagcgagcaga 180
    |||||||
Db 131 atcgagcctacgcgcttccctcgctggtgtgttcgacaaacaggaagcgagcaga 190
QY 181 gtccaaacgctggtgagagcgctccgagcagatcctcgacatccaccgagcagctg 240
    |||||||
Db 191 gtccaaacgctggtgagagcgctccgagcagatcctcgacatccaccgagcagctg 250
QY 241 atcacgcggtgagagcgctccgacatccgacgcgagatggcgagcgcttaac 300
    |||||||
Db 251 atcacgcggtgagagcgctccgacatccgacgcgagatggcgagcgcttaac 310
QY 301 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 360
    |||||||
Db 311 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 370
QY 361 acagggaacgtgacatctgcccgaaggaccccgcc 396
    |||||||
Db 371 acagggaacgtgacatctgcccgaaggaccccgcc 406

RESULT 6
BD006325 447 bp DNA linear PAT 31-JAN-2002
LOCUS BD006325
DEFINITION Compounds and methods for diagnosis of Tuberculosis.
ACCESSION BD006325
VERSION BD006325.1 GI:18634696
KEYWORDS JP 2001500383-A/4.
SOURCE unidentified.

```

```

ORGANISM unidentified
REFERENCE 1 (bases 1 to 447)
AUTHORS Reed,S.G., Skelky,Y.A.W., Dillon,D.C., Neto,A.C., Houghton,R.,
          Vedvik,T.S., Twardzik,D.R. and Lodes,M.J.
TITLE Compounds and methods for diagnosis of Tuberculosis
JOURNAL Patent: JP 2001500383-A 4 16-JAN-2001;
CORIXA CORP
COMMENT OS Unidentified
        PN JP 2001500383-A/4
        PD 16-JAN-2001
        PE 07-OCT-1997 JP 1998518432
        PR 11-OCT-1996 US 08/729622,13-MAR-1997 US 08/818111 PI
          STEVEN G REED,YASIR A W SKEIKY,DAVIN C DILLON, PI ANTONIO CAMPOS
          NETO,
          PI RAYMOND HOUGHTON,THOMAS S VEDVIK,DANIEL R TWARZIK, PI
          MICHAEL J LODES
PC C12N15/31,C07K14/35,C07K16/12,C12Q1/68,C12N15/62,G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..447
   /organism='Unidentified'.
FEATURES Location/Qualifiers
source 1..447
BASE COUNT 79 a 146 c 149 g 72 t 1 others
ORIGIN

Query Match 99.6%; Score 394.4; DB 6; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.6e-55;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcgctcgatatacttccagctgctccagggcgagggatcgccattccgac 60
    |||||||
Db 11 acggcgcgctcgatatacttccagctgctccagggcgagggatcgccattccgac 70
QY 61 gggcagcgatgacatcgccgagcagatccgacatcggtggggggtcaccacgcttcat 120
    |||||||
Db 71 gggcagcgatgacatcgccgagcagatccgacatcggtggggggtcaccacgcttcat 130
QY 121 atcgagcctacgcgcttccctcgctggtgtgttcgacaaacaggaagcgagcaga 180
    |||||||
Db 131 atcgagcctacgcgcttccctcgctggtgtgttcgacaaacaggaagcgagcaga 190
QY 181 gtccaaacgctggtgagagcgctccgagcagatcctcgacatccaccgagcagctg 240
    |||||||
Db 191 gtccaaacgctggtgagagcgctccgagcagatcctcgacatccaccgagcagctg 250
QY 241 atcacgcggtgagagcgctccgacatccgacgcgagatggcgagcgcttaac 300
    |||||||
Db 251 atcacgcggtgagagcgctccgacatccgacgcgagatggcgagcgcttaac 310
QY 301 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 360
    |||||||
Db 311 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 370
QY 361 acagggaacgtgacatctgcccgaaggaccccgcc 396
    |||||||
Db 371 acagggaacgtgacatctgcccgaaggaccccgcc 406

RESULT 7
BD006445 447 bp DNA linear PAT 31-JAN-2002
LOCUS BD006445
DEFINITION Compounds and methods for immunotherapy and diagnosis of
          Tuberculosis.
ACCESSION BD006445
VERSION BD006445.1 GI:18634816
KEYWORDS JP 2001501832-A/4.
SOURCE unidentified.

```


NETO,
PI RAYMOND HOUGHTON, THOMAS S VEDVICK, DANIEL R TWARDZIK, PI
MICHAEL J LODES
PC C12N15/31, C07K14/35, C07K16/12, C12Q1/68, C12N15/62, G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1.1872
Location/Qualifiers
/organism='Unidentified'.
/db_xref='taxon:32644'

BASE COUNT 318 a 616 c 604 g 332 t 2 others
ORIGIN

Query Match 98.8%; Score 391.2; DB 6; Length 1872;
Best Local Similarity 99.2%; Pred. No. 4, 2e-55;
Matches 393; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 acggccgctgcgataactccagctgtccagagtgaggagattccgcatccgcatc 60
|||||
DB 758 ACGGCCGCTGCCATTAATCTCCAGCTGTCCAGAGTGAGGAGATTCGCATTCGATC 817
|||||
QY 61 gggcagcagatgagcagtcgagcagatccgagtggtggggtgacccacgcttcat 120
|||||
DB 818 GGGCAGGCGATGCGATCCGCGGCAATCCGATCGGAGGAGTCCACCCGCTTCAT 877
|||||
QY 121 atcgagctacgagctccctccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
|||||
DB 878 ATCGGCGCTACCGCTCTCCGCTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 937
|||||
QY 181 gtccaacgctgt 240
|||||
DB 938 GTCCAAACGCTGT 997
|||||
QY 241 ataccagcgt 300
|||||
DB 998 ATACCGCGCTGT 1057
|||||
QY 301 gggcagcagatgagcagtcgagcagatccgagtggtggggtgacccacgcttcat 360
|||||
DB 1058 GGGCAGGCGATGCGATCCGCGGCAATCCGATCGGAGGAGTTCGCATTCGATC 1117
|||||
QY 361 acagggaacgtgacatgtgacgagagagagagagagagagagagagagagagag 396
|||||
DB 1118 ACAGGGGAACGTGACATTGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1153
|||||

RESULT 10
BD006458 1872 bp DNA linear PAT 31-JAN-2002
LOCUS
DEFINITION
Tuberculosis.
ACCESSION
BD006458
VERSION
BD006458.1 GI:18634829
KEYWORDS
JP 2001501832-A/17.
SOURCE
unclassified.
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 1872)
AUTHORS
Reed, S.G., Skeiky, Y.A.W., Dillon, D.C., Neto, A.C., Houghton, R.,
Vedvick, T.S., Twardzik, D.R., and Lodes, M.J.
TITLE
Compounds and methods for immunotherapy and diagnosis of
PATENT: JP 2001501832-A 17 13-FEB-2001;
JOURNAL
CORIXA CORP
COMMENT
OS Unidentified
PN JP 2001501832-A/17
PD 13-FEB-2001
PF 07-OCT-1997 JP 1998518456
PR 11-OCT-1996 US 08/730510, 13-MAR-1997 US 08/818112 PT
STEVEN G REED, YASIR A W SKEIKY, DAVIN C DILLON, PI ANTONIO CAMPOS

NETO,
PI RAYMOND HOUGHTON, THOMAS S VEDVICK, DANIEL R TWARDZIK, PI
MICHAEL J LODES
PC C12N15/31, C07K14/35, A61K39/04, A61K48/00, A61K49/00, C12N15/62,
PC C07K19/00,
PC G01N33/50, G01N33/60, G01N33/569, C12N1/19, C12N1/20, C12N1/21, PC
C12N5/10//
PC (C12N1/21, C12R1:19)
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1.1872
Location/Qualifiers
/organism='Unidentified'.
/db_xref='taxon:32644'

BASE COUNT 318 a 616 c 604 g 332 t 2 others
ORIGIN

Query Match 98.8%; Score 391.2; DB 6; Length 1872;
Best Local Similarity 99.2%; Pred. No. 4, 2e-55;
Matches 393; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 acggccgctgcgataactccagctgtccagagtgaggagattccgcatccgcatc 60
|||||
DB 758 ACGGCCGCTGCCATTAATCTCCAGCTGTCCAGAGTGAGGAGATTCGCATTCGATC 817
|||||
QY 61 gggcagcagatgagcagtcgagcagatccgagtggtggggtgacccacgcttcat 120
|||||
DB 818 GGGCAGGCGATGCGATCCGCGGCAATCCGATCGGAGGAGTCCACCCGCTTCAT 877
|||||
QY 121 atcgagctacgagctccctccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
|||||
DB 878 ATCGGCGCTACCGCTCTCCGCTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 937
|||||
QY 181 gtccaacgctgt 240
|||||
DB 938 GTCCAAACGCTGT 997
|||||
QY 241 ataccagcgt 300
|||||
DB 998 ATACCGCGCTGT 1057
|||||
QY 301 gggcagcagatgagcagtcgagcagatccgagtggtggggtgacccacgcttcat 360
|||||
DB 1058 GGGCAGGCGATGCGATCCGCGGCAATCCGATCGGAGGAGTTCGCATTCGATC 1117
|||||
QY 361 acagggaacgtgacatgtgacgagagagagagagagagagagagagagagagag 396
|||||
DB 1118 ACAGGGGAACGTGACATTGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1153
|||||

RESULT 11
AX201049 675 bp DNA linear PAT 29-AUG-2001
LOCUS
DEFINITION
Sequence 679 from Patent WO0151633.
ACCESSION
AX201049
VERSION
AX201049.1 GI:15390857
KEYWORDS
human.
SOURCE
ORGANISM
human.
REFERENCE
1 (bases 1 to 675)
AUTHORS
Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Reed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
Stolk, J.A., Skeiky, Y.A., Wang, A., and Meagher, M.J.
TITLE
Compositions and methods for the therapy and diagnosis of prostate
cancer
PATENT: WO 0151633-A 679 19-JUL-2001;
JOURNAL
CORIXA CORPORATION (US)


```

QY 241 ataccgaggtcagcgctccatcaactcggccacgcgagtcgagcgagctaac 300
    |||||||
DB 250 ATACCGCGGTGACGCGGCTCGATCAATCGGCGCATGTGGGAGCGGCTTAAC 309
    |||||||
QY 301 gggcatcaccggtgagctcattcgtgagcttgcaaacgaagtcggcgagcgct 360
    |||||||
DB 310 GGGCATCATCCCGGTGACGTCATCTCGTGACCTGGCAAAAGTGGGCGGCGCGT 369
    |||||||
QY 361 acaggaagctgacattgagcgagggagcccgcc 396
    |||||||
DB 370 ACAGGGAACCTGACATTGGCCGAGGAGACCCCGGCC 405
    |||||||

RESULT 14
LOCUS AX201061 915 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 691 from Patent WO0151633.
ACCESSION AX201061
VERSION AX201061.1 GI:15390868
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 915)
Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Stolk,J.A., Skeiky,Y.A., Wang,A., and Meagher,M.J.
Compositions and methods for the therapy and diagnosis of prostate
cancer
Patent: WO 0151633-A 691 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source location/Qualifiers
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 167 a 280 c 284 g 184 t
ORIGIN
Query Match 89.4%; Score 354; DB 6; Length 915;
Best Local Similarity 95.7%; Pred. No. 5.7e-49;
Matches 379; Conservative 0; Mismatches 5; Indels 12; Gaps 1;

QY 1 acggcgctgcgataactccagctgctccaggtgagcgagatccgcatccgac 60
    |||||||
DB 22 ACGGCGGCTCGCATTAATCTCCAGCTGTCCAGGTTGGCAGGATTCGCATTCGATC 81
    |||||||
QY 61 gggcagcgatgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 120
    |||||||
DB 82 GGGCAGGCGATGAGTCGAGTCGCGGCGCAGATC-----AAGCTTCCACCGCTTCAT 129
    |||||||
QY 121 atcgagctaccgctcctccgctgctggtgtgtgtgcagacaacagcgagcgag 180
    |||||||
DB 130 ATCGGCGCTACCGCTCTCTCGGCTTGGTGTGTGTGCACAAACGCGACGCGACGA 189
    |||||||
QY 181 gtccaaagctgctgagcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 240
    |||||||
DB 190 GTCCAAAGCTGTGTCGAGCGCTCCGCGGCAAGTCTCGCATCTCCACCGGAGCTG 249
    |||||||
QY 241 ataccgaggtcagcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 300
    |||||||
DB 250 ATACCGCGGTGACGCGGCTCGATCAATCGGCGCATGTGGGAGCGGCTTAAC 309
    |||||||
QY 301 gggcatcaccggtgagctcattcgtgagcttgcaaacgaagtcggcgagcgct 360
    |||||||
DB 310 GGGCATCATCCCGGTGACGTCATCTCGTGACCTGGCAAAAGTGGGCGGCGCGT 369
    |||||||
QY 361 acaggaagctgacattgagcgagggagcccgcc 396
    |||||||
DB 370 ACAGGGAACCTGACATTGGCCGAGGAGACCCCGGCC 405
    |||||||

```

```

RESULT 15
LOCUS AX267860 915 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 834 from Patent WO0173032.
ACCESSION AX267860
VERSION AX267860.1 GI:16516503
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
and Henderson,R.A.
Compositions and methods for the therapy and diagnosis of prostate
cancer
Patent: WO 0173032-A 834 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source location/Qualifiers
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 167 a 280 c 284 g 184 t
ORIGIN
Query Match 89.4%; Score 354; DB 6; Length 915;
Best Local Similarity 95.7%; Pred. No. 5.7e-49;
Matches 379; Conservative 0; Mismatches 5; Indels 12; Gaps 1;

QY 1 acggcgctgcgataactccagctgctccaggtgagcgagatccgcatccgac 60
    |||||||
DB 22 ACGGCGGCTCGCATTAATCTCCAGCTGTCCAGGTTGGCAGGATTCGCATTCGATC 81
    |||||||
QY 61 gggcagcgatgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 120
    |||||||
DB 82 GGGCAGGCGATGAGTCGAGTCGCGGCGCAGATC-----AAGCTTCCACCGCTTCAT 129
    |||||||
QY 121 atcgagctaccgctcctccgctgctggtgtgtgtgcagacaacagcgagcgag 180
    |||||||
DB 130 ATCGGCGCTACCGCTCTCTCGGCTTGGTGTGTGTGCACAAACGCGACGCGACGA 189
    |||||||
QY 181 gtccaaagctgctgagcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 240
    |||||||
DB 190 GTCCAAAGCTGTGTCGAGCGCTCCGCGGCAAGTCTCGCATCTCCACCGGAGCTG 249
    |||||||
QY 241 ataccgaggtcagcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 300
    |||||||
DB 250 ATACCGCGGTGACGCGGCTCGATCAATCGGCGCATGTGGGAGCGGCTTAAC 309
    |||||||
QY 301 gggcatcaccggtgagctcattcgtgagcttgcaaacgaagtcggcgagcgct 360
    |||||||
DB 310 GGGCATCATCCCGGTGACGTCATCTCGTGACCTGGCAAAAGTGGGCGGCGCGT 369
    |||||||
QY 361 acaggaagctgacattgagcgagggagcccgcc 396
    |||||||
DB 370 ACAGGGAACCTGACATTGGCCGAGGAGACCCCGGCC 405
    |||||||

```

Search completed: August 6, 2002, 11:05:22
 Job time: 7707 sec

KM Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KM hybridisation; detection; vaccine; immunisation; infection; ss.
 XX Mycobacterium sp.
 OS
 PN W09909186-A2. *Compiling*
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR01813.
 XX
 PR 11-SEP-1997; 97FR-0011325.
 PR 14-AUG-1997; 97FR-0010404.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Glucquel B, Lim EM, Pelletier V, Portnoi D, Goguet de la Salmoniere Y;
 PI Guigneno A;
 XX
 DR WPI; 1999-181045/15.
 DR P-PSDB; AAY04830.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 PS
 PS Claim 22; Fig 50F; 309pp; French.
 XX
 CC Sequences AAX34001-X34352 represent nucleic acids encoding secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC nucleotide sequences can be used as primers and probes for methods
 CC for detecting and identifying mycobacteria, especially belonging to
 CC the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 XX
 SO Sequence 1143 BP; 189 A; 373 C; 395 G; 186 T; 0 other;

Query Match 100.0%; Score 396; DB 20; Length 1143;
 Best Local Similarity 100.0%; Pred. No. 2.1e-85;
 Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acggcgcgctccgatactccagctccagagtgaggagagatccgcatccgac 60
 DB 745 acggcgcgctccgatactccagctccagagtgaggagagatccgcatccgac 804
 QY 61 gggcaggcgatgagatcgcgagcgagatcgatcggtgggggtcaaccacgttcat 120
 DB 805 gggcaggcgatgagatcgcgagcgagatcgatcggtgggggtcaaccacgttcat 864
 QY 121 atcgggctaccgctctccgcttggtgtgtgtcgaacaacaagcagcagcagca 180
 DB 865 atcgggctaccgctctccgcttggtgtgtgtcgaacaacaagcagcagcagca 924
 QY 181 gtccaaagcggt 240
 DB 925 gtccaaagcggt 984
 QY 241 atcacacgctgctgacgagcgtccgatacaactcgcgcacacgagatgagcagcgttacc 300
 DB 985 atcacacgctgctgacgagcgtccgatacaactcgcgcacacgagatgagcagcgttacc 1044
 QY 301 gggcattatcccggtgagatgcatctcgtgtgacttgcaaaccaatcggcgagcagcgt 360
 DB 1045 gggcattatcccggtgagatgcatctcgtgtgacttgcaaaccaatcggcgagcagcgt 1104
 QY 361 acaggaagcgtgacatgagcagagagaccccgagc 396
 DB 1105 acaggaagcgtgacatgagcagagagaccccgagc 1140

RESULT 4
 AAZ20194

ID AAZ20194 standard; DNA; 2287 BP.
 XX
 AC AAZ20194;
 XX
 DE 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein Mtub32A DNA.
 XX
 KM Tuberculosis; antigen; fusion protein; Mtub32A; Ra12; Tbh9; Ra35;
 KM diagnosis; therapy; vaccine; immunogen; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PH Key Location/Qualifiers
 FT CDS 42..2231
 FT /tag= a
 XX
 PN W09951748-A2.
 XX
 PD 14-OCT-1999.
 XX
 PE 07-APR-1999; 99NO-US07717.
 XX
 PR 07-APR-1998; 98US-0056556.
 PR 30-DEC-1998; 98US-0223040.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAM, Alderson M, Campos-Neto A;
 XX
 DR WPI; 1999-601610/51.
 DR P-PSDB; AAY32059.
 XX
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis -
 XX
 PS Example; Fig 1A-B; 83pp; English.
 XX
 CC This DNA sequence includes a coding region for a recombinant
 CC Mycobacterium tuberculosis tri-antigen fusion protein (see AAY32059),
 CC termed Mtub32A, composed of the antigens Ra12, Tbh9 and Ra35. The
 CC DNA is useful for the recombinant production of the fusion protein.
 CC Coding sequences for the antigens were modified by PCR in order
 CC to facilitate their fusion and subsequent expression of the fusion
 CC protein. 3 Coding sequences for Ra12, Tbh9 and Ra35 were ligated
 CC to encode Mtub32A. The invention provides fusion proteins (see
 CC AAY32059-71) containing at least 2 M. tuberculosis antigens. The new
 CC fusion proteins and polynucleotides encoding them are useful as
 CC vaccines for preventing tuberculosis (claimed), for diagnosis (via
 CC in vitro assays or intradermal skin tests for detection of anti-M.
 CC tuberculosis antibodies), monitoring of disease progression, and
 CC treatment of tuberculosis. They are more effective immunogens than
 CC mixtures of the individual protein components.
 XX

SO Sequence 2287 BP; 374 A; 723 C; 805 G; 382 T; 3 other;

Query Match 100.0%; Score 396; DB 20; Length 2287;
 Best Local Similarity 100.0%; Pred. No. 2.2e-85;
 Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acggcgcgctccgatactccagctccagagtgaggagagatccgcatccgac 60
 DB 63 acggcgcgctccgatactccagctccagagtgaggagagatccgcatccgac 122
 QY 61 gggcaggcgatgagatcgcgagcgagatcgatcggtgggggtcaaccacgttcat 120
 DB 123 gggcaggcgatgagatcgcgagcgagatcgatcggtgggggtcaaccacgttcat 182
 QY 121 atcgggctaccgctctccgcttggtgtgtgtcgaacaacaagcagcagcagca 180
 DB 183 atcgggctaccgctctccgcttggtgtgtgtcgaacaacaagcagcagcagca 242

QY 181 gtccacgcgtgtgctggagcgctccgcgcgaagtcttcgcatccacggcgagctg 240
DB 243 gtccacgcgtgtgctggagcgctccgcgcgaagtcttcgcatccacggcgagctg 302
QY 241 atcacgcggttcgacgcgctccgcgcgaagtcttcgcatccacggcgagctg 300
DB 303 atcacgcggttcgacgcgctccgcgcgaagtcttcgcatccacggcgagctg 362
QY 301 gggcgcgcgttcgacgcgctccgcgcgaagtcttcgcatccacggcgagctg 360
DB 363 gggcgcgcgttcgacgcgctccgcgcgaagtcttcgcatccacggcgagctg 422
QY 361 acagggacgtgacgttcgacgcgctccgcgcgaagtcttcgcatccacggcgagctg 396
DB 423 acagggacgtgacgttcgacgcgctccgcgcgaagtcttcgcatccacggcgagctg 458

RESULT 5

AA199683
ID AA199683 standard; DNA: 4403765 BP.

AC AA199683;

DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome; variation; epidemiology; patient treatment; epidemic monitoring; ds.

OS Mycobacterium tuberculosis.

PN US6294328-B1. *X. ¹ ² ³ ⁴ ⁵ ⁶ ⁷ ⁸ ⁹ ¹⁰ ¹¹ ¹² ¹³ ¹⁴ ¹⁵ ¹⁶ ¹⁷ ¹⁸ ¹⁹ ²⁰ ²¹ ²² ²³ ²⁴ ²⁵ ²⁶ ²⁷ ²⁸ ²⁹ ³⁰ ³¹ ³² ³³ ³⁴ ³⁵ ³⁶ ³⁷ ³⁸ ³⁹ ⁴⁰ ⁴¹ ⁴² ⁴³ ⁴⁴ ⁴⁵ ⁴⁶ ⁴⁷ ⁴⁸ ⁴⁹ ⁵⁰ ⁵¹ ⁵² ⁵³ ⁵⁴ ⁵⁵ ⁵⁶ ⁵⁷ ⁵⁸ ⁵⁹ ⁶⁰ ⁶¹ ⁶² ⁶³ ⁶⁴ ⁶⁵ ⁶⁶ ⁶⁷ ⁶⁸ ⁶⁹ ⁷⁰ ⁷¹ ⁷² ⁷³ ⁷⁴ ⁷⁵ ⁷⁶ ⁷⁷ ⁷⁸ ⁷⁹ ⁸⁰ ⁸¹ ⁸² ⁸³ ⁸⁴ ⁸⁵ ⁸⁶ ⁸⁷ ⁸⁸ ⁸⁹ ⁹⁰ ⁹¹ ⁹² ⁹³ ⁹⁴ ⁹⁵ ⁹⁶ ⁹⁷ ⁹⁸ ⁹⁹ ¹⁰⁰ ¹⁰¹ ¹⁰² ¹⁰³ ¹⁰⁴ ¹⁰⁵ ¹⁰⁶ ¹⁰⁷ ¹⁰⁸ ¹⁰⁹ ¹¹⁰ ¹¹¹ ¹¹² ¹¹³ ¹¹⁴ ¹¹⁵ ¹¹⁶ ¹¹⁷ ¹¹⁸ ¹¹⁹ ¹²⁰ ¹²¹ ¹²² ¹²³ ¹²⁴ ¹²⁵ ¹²⁶ ¹²⁷ ¹²⁸ ¹²⁹ ¹³⁰ ¹³¹ ¹³² ¹³³ ¹³⁴ ¹³⁵ ¹³⁶ ¹³⁷ ¹³⁸ ¹³⁹ ¹⁴⁰ ¹⁴¹ ¹⁴² ¹⁴³ ¹⁴⁴ ¹⁴⁵ ¹⁴⁶ ¹⁴⁷ ¹⁴⁸ ¹⁴⁹ ¹⁵⁰ ¹⁵¹ ¹⁵² ¹⁵³ ¹⁵⁴ ¹⁵⁵ ¹⁵⁶ ¹⁵⁷ ¹⁵⁸ ¹⁵⁹ ¹⁶⁰ ¹⁶¹ ¹⁶² ¹⁶³ ¹⁶⁴ ¹⁶⁵ ¹⁶⁶ ¹⁶⁷ ¹⁶⁸ ¹⁶⁹ ¹⁷⁰ ¹⁷¹ ¹⁷² ¹⁷³ ¹⁷⁴ ¹⁷⁵ ¹⁷⁶ ¹⁷⁷ ¹⁷⁸ ¹⁷⁹ ¹⁸⁰ ¹⁸¹ ¹⁸² ¹⁸³ ¹⁸⁴ ¹⁸⁵ ¹⁸⁶ ¹⁸⁷ ¹⁸⁸ ¹⁸⁹ ¹⁹⁰ ¹⁹¹ ¹⁹² ¹⁹³ ¹⁹⁴ ¹⁹⁵ ¹⁹⁶ ¹⁹⁷ ¹⁹⁸ ¹⁹⁹ ²⁰⁰ ²⁰¹ ²⁰² ²⁰³ ²⁰⁴ ²⁰⁵ ²⁰⁶ ²⁰⁷ ²⁰⁸ ²⁰⁹ ²¹⁰ ²¹¹ ²¹² ²¹³ ²¹⁴ ²¹⁵ ²¹⁶ ²¹⁷ ²¹⁸ ²¹⁹ ²²⁰ ²²¹ ²²² ²²³ ²²⁴ ²²⁵ ²²⁶ ²²⁷ ²²⁸ ²²⁹ ²³⁰ ²³¹ ²³² ²³³ ²³⁴ ²³⁵ ²³⁶ ²³⁷ ²³⁸ ²³⁹ ²⁴⁰ ²⁴¹ ²⁴² ²⁴³ ²⁴⁴ ²⁴⁵ ²⁴⁶ ²⁴⁷ ²⁴⁸ ²⁴⁹ ²⁵⁰ ²⁵¹ ²⁵² ²⁵³ ²⁵⁴ ²⁵⁵ ²⁵⁶ ²⁵⁷ ²⁵⁸ ²⁵⁹ ²⁶⁰ ²⁶¹ ²⁶² ²⁶³ ²⁶⁴ ²⁶⁵ ²⁶⁶ ²⁶⁷ ²⁶⁸ ²⁶⁹ ²⁷⁰ ²⁷¹ ²⁷² ²⁷³ ²⁷⁴ ²⁷⁵ ²⁷⁶ ²⁷⁷ ²⁷⁸ ²⁷⁹ ²⁸⁰ ²⁸¹ ²⁸² ²⁸³ ²⁸⁴ ²⁸⁵ ²⁸⁶ ²⁸⁷ ²⁸⁸ ²⁸⁹ ²⁹⁰ ²⁹¹ ²⁹² ²⁹³ ²⁹⁴ ²⁹⁵ ²⁹⁶ ²⁹⁷ ²⁹⁸ ²⁹⁹ ³⁰⁰ ³⁰¹ ³⁰² ³⁰³ ³⁰⁴ ³⁰⁵ ³⁰⁶ ³⁰⁷ ³⁰⁸ ³⁰⁹ ³¹⁰ ³¹¹ ³¹² ³¹³ ³¹⁴ ³¹⁵ ³¹⁶ ³¹⁷ ³¹⁸ ³¹⁹ ³²⁰ ³²¹ ³²² ³²³ ³²⁴ ³²⁵ ³²⁶ ³²⁷ ³²⁸ ³²⁹ ³³⁰ ³³¹ ³³² ³³³ ³³⁴ ³³⁵ ³³⁶ ³³⁷ ³³⁸ ³³⁹ ³⁴⁰ ³⁴¹ ³⁴² ³⁴³ ³⁴⁴ ³⁴⁵ ³⁴⁶ ³⁴⁷ ³⁴⁸ ³⁴⁹ ³⁵⁰ ³⁵¹ ³⁵² ³⁵³ ³⁵⁴ ³⁵⁵ ³⁵⁶ ³⁵⁷ ³⁵⁸ ³⁵⁹ ³⁶⁰ ³⁶¹ ³⁶² ³⁶³ ³⁶⁴ ³⁶⁵ ³⁶⁶ ³⁶⁷ ³⁶⁸ ³⁶⁹ ³⁷⁰ ³⁷¹ ³⁷² ³⁷³ ³⁷⁴ ³⁷⁵ ³⁷⁶ ³⁷⁷ ³⁷⁸ ³⁷⁹ ³⁸⁰ ³⁸¹ ³⁸² ³⁸³ ³⁸⁴ ³⁸⁵ ³⁸⁶ ³⁸⁷ ³⁸⁸ ³⁸⁹ ³⁹⁰ ³⁹¹ ³⁹² ³⁹³ ³⁹⁴ ³⁹⁵ ³⁹⁶ ³⁹⁷ ³⁹⁸ ³⁹⁹ ⁴⁰⁰ ⁴⁰¹ ⁴⁰² ⁴⁰³ ⁴⁰⁴ ⁴⁰⁵ ⁴⁰⁶ ⁴⁰⁷ ⁴⁰⁸ ⁴⁰⁹ ⁴¹⁰ ⁴¹¹ ⁴¹² ⁴¹³ ⁴¹⁴ ⁴¹⁵ ⁴¹⁶ ⁴¹⁷ ⁴¹⁸ ⁴¹⁹ ⁴²⁰ ⁴²¹ ⁴²² ⁴²³ ⁴²⁴ ⁴²⁵ ⁴²⁶ ⁴²⁷ ⁴²⁸ ⁴²⁹ ⁴³⁰ ⁴³¹ ⁴³² ⁴³³ ⁴³⁴ ⁴³⁵ ⁴³⁶ ⁴³⁷ ⁴³⁸ ⁴³⁹ ⁴⁴⁰ ⁴⁴¹ ⁴⁴² ⁴⁴³ ⁴⁴⁴ ⁴⁴⁵ ⁴⁴⁶ ⁴⁴⁷ ⁴⁴⁸ ⁴⁴⁹ ⁴⁵⁰ ⁴⁵¹ ⁴⁵² ⁴⁵³ ⁴⁵⁴ ⁴⁵⁵ ⁴⁵⁶ ⁴⁵⁷ ⁴⁵⁸ ⁴⁵⁹ ⁴⁶⁰ ⁴⁶¹ ⁴⁶² ⁴⁶³ ⁴⁶⁴ ⁴⁶⁵ ⁴⁶⁶ ⁴⁶⁷ ⁴⁶⁸ ⁴⁶⁹ ⁴⁷⁰ ⁴⁷¹ ⁴⁷² ⁴⁷³ ⁴⁷⁴ ⁴⁷⁵ ⁴⁷⁶ ⁴⁷⁷ ⁴⁷⁸ ⁴⁷⁹ ⁴⁸⁰ ⁴⁸¹ ⁴⁸² ⁴⁸³ ⁴⁸⁴ ⁴⁸⁵ ⁴⁸⁶ ⁴⁸⁷ ⁴⁸⁸ ⁴⁸⁹ ⁴⁹⁰ ⁴⁹¹ ⁴⁹² ⁴⁹³ ⁴⁹⁴ ⁴⁹⁵ ⁴⁹⁶ ⁴⁹⁷ ⁴⁹⁸ ⁴⁹⁹ ⁵⁰⁰ ⁵⁰¹ ⁵⁰² ⁵⁰³ ⁵⁰⁴ ⁵⁰⁵ ⁵⁰⁶ ⁵⁰⁷ ⁵⁰⁸ ⁵⁰⁹ ⁵¹⁰ ⁵¹¹ ⁵¹² ⁵¹³ ⁵¹⁴ ⁵¹⁵ ⁵¹⁶ ⁵¹⁷ ⁵¹⁸ ⁵¹⁹ ⁵²⁰ ⁵²¹ ⁵²² ⁵²³ ⁵²⁴ ⁵²⁵ ⁵²⁶ ⁵²⁷ ⁵²⁸ ⁵²⁹ ⁵³⁰ ⁵³¹ ⁵³² ⁵³³ ⁵³⁴ ⁵³⁵ ⁵³⁶ ⁵³⁷ ⁵³⁸ ⁵³⁹ ⁵⁴⁰ ⁵⁴¹ ⁵⁴² ⁵⁴³ ⁵⁴⁴ ⁵⁴⁵ ⁵⁴⁶ ⁵⁴⁷ ⁵⁴⁸ ⁵⁴⁹ ⁵⁵⁰ ⁵⁵¹ ⁵⁵² ⁵⁵³ ⁵⁵⁴ ⁵⁵⁵ ⁵⁵⁶ ⁵⁵⁷ ⁵⁵⁸ ⁵⁵⁹ ⁵⁶⁰ ⁵⁶¹ ⁵⁶² ⁵⁶³ ⁵⁶⁴ ⁵⁶⁵ ⁵⁶⁶ ⁵⁶⁷ ⁵⁶⁸ ⁵⁶⁹ ⁵⁷⁰ ⁵⁷¹ ⁵⁷² ⁵⁷³ ⁵⁷⁴ ⁵⁷⁵ ⁵⁷⁶ ⁵⁷⁷ ⁵⁷⁸ ⁵⁷⁹ ⁵⁸⁰ ⁵⁸¹ ⁵⁸² ⁵⁸³ ⁵⁸⁴ ⁵⁸⁵ ⁵⁸⁶ ⁵⁸⁷ ⁵⁸⁸ ⁵⁸⁹ ⁵⁹⁰ ⁵⁹¹ ⁵⁹² ⁵⁹³ ⁵⁹⁴ ⁵⁹⁵ ⁵⁹⁶ ⁵⁹⁷ ⁵⁹⁸ ⁵⁹⁹ ⁶⁰⁰ ⁶⁰¹ ⁶⁰² ⁶⁰³ ⁶⁰⁴ ⁶⁰⁵ ⁶⁰⁶ ⁶⁰⁷ ⁶⁰⁸ ⁶⁰⁹ ⁶¹⁰ ⁶¹¹ ⁶¹² ⁶¹³ ⁶¹⁴ ⁶¹⁵ ⁶¹⁶ ⁶¹⁷ ⁶¹⁸ ⁶¹⁹ ⁶²⁰ ⁶²¹ ⁶²² ⁶²³ ⁶²⁴ ⁶²⁵ ⁶²⁶ ⁶²⁷ ⁶²⁸ ⁶²⁹ ⁶³⁰ ⁶³¹ ⁶³² ⁶³³ ⁶³⁴ ⁶³⁵ ⁶³⁶ ⁶³⁷ ⁶³⁸ ⁶³⁹ ⁶⁴⁰ ⁶⁴¹ ⁶⁴² ⁶⁴³ ⁶⁴⁴ ⁶⁴⁵ ⁶⁴⁶ ⁶⁴⁷ ⁶⁴⁸ ⁶⁴⁹ ⁶⁵⁰ ⁶⁵¹ ⁶⁵² ⁶⁵³ ⁶⁵⁴ ⁶⁵⁵ ⁶⁵⁶ ⁶⁵⁷ ⁶⁵⁸ ⁶⁵⁹ ⁶⁶⁰ ⁶⁶¹ ⁶⁶² ⁶⁶³ ⁶⁶⁴ ⁶⁶⁵ ⁶⁶⁶ ⁶⁶⁷ ⁶⁶⁸ ⁶⁶⁹ ⁶⁷⁰ ⁶⁷¹ ⁶⁷² ⁶⁷³ ⁶⁷⁴ ⁶⁷⁵ ⁶⁷⁶ ⁶⁷⁷ ⁶⁷⁸ ⁶⁷⁹ ⁶⁸⁰ ⁶⁸¹ ⁶⁸² ⁶⁸³ ⁶⁸⁴ ⁶⁸⁵ ⁶⁸⁶ ⁶⁸⁷ ⁶⁸⁸ ⁶⁸⁹ ⁶⁹⁰ ⁶⁹¹ ⁶⁹² ⁶⁹³ ⁶⁹⁴ ⁶⁹⁵ ⁶⁹⁶ ⁶⁹⁷ ⁶⁹⁸ ⁶⁹⁹ ⁷⁰⁰ ⁷⁰¹ ⁷⁰² ⁷⁰³ ⁷⁰⁴ ⁷⁰⁵ ⁷⁰⁶ ⁷⁰⁷ ⁷⁰⁸ ⁷⁰⁹ ⁷¹⁰ ⁷¹¹ ⁷¹² ⁷¹³ ⁷¹⁴ ⁷¹⁵ ⁷¹⁶ ⁷¹⁷ ⁷¹⁸ ⁷¹⁹ ⁷²⁰ ⁷²¹ ⁷²² ⁷²³ ⁷²⁴ ⁷²⁵ ⁷²⁶ ⁷²⁷ ⁷²⁸ ⁷²⁹ ⁷³⁰ ⁷³¹ ⁷³² ⁷³³ ⁷³⁴ ⁷³⁵ ⁷³⁶ ⁷³⁷ ⁷³⁸ ⁷³⁹ ⁷⁴⁰ ⁷⁴¹ ⁷⁴² ⁷⁴³ ⁷⁴⁴ ⁷⁴⁵ ⁷⁴⁶ ⁷⁴⁷ ⁷⁴⁸ ⁷⁴⁹ ⁷⁵⁰ ⁷⁵¹ ⁷⁵² ⁷⁵³ ⁷⁵⁴ ⁷⁵⁵ ⁷⁵⁶ ⁷⁵⁷ ⁷⁵⁸ ⁷⁵⁹ ⁷⁶⁰ ⁷⁶¹ ⁷⁶² ⁷⁶³ ⁷⁶⁴ ⁷⁶⁵ ⁷⁶⁶ ⁷⁶⁷ ⁷⁶⁸ ⁷⁶⁹ ⁷⁷⁰ ⁷⁷¹ ⁷⁷² ⁷⁷³ ⁷⁷⁴ ⁷⁷⁵ ⁷⁷⁶ ⁷⁷⁷ ⁷⁷⁸ ⁷⁷⁹ ⁷⁸⁰ ⁷⁸¹ ⁷⁸² ⁷⁸³ ⁷⁸⁴ ⁷⁸⁵ ⁷⁸⁶ ⁷⁸⁷ ⁷⁸⁸ ⁷⁸⁹ ⁷⁹⁰ ⁷⁹¹ ⁷⁹² ⁷⁹³ ⁷⁹⁴ ⁷⁹⁵ ⁷⁹⁶ ⁷⁹⁷ ⁷⁹⁸ ⁷⁹⁹ ⁸⁰⁰ ⁸⁰¹ ⁸⁰² ⁸⁰³ ⁸⁰⁴ ⁸⁰⁵ ⁸⁰⁶ ⁸⁰⁷ ⁸⁰⁸ ⁸⁰⁹ ⁸¹⁰ ⁸¹¹ ⁸¹² ⁸¹³ ⁸¹⁴ ⁸¹⁵ ⁸¹⁶ ⁸¹⁷ ⁸¹⁸ ⁸¹⁹ ⁸²⁰ ⁸²¹ ⁸²² ⁸²³ ⁸²⁴ ⁸²⁵ ⁸²⁶ ⁸²⁷ ⁸²⁸ ⁸²⁹ ⁸³⁰ ⁸³¹ ⁸³² ⁸³³ ⁸³⁴ ⁸³⁵ ⁸³⁶ ⁸³⁷ ⁸³⁸ ⁸³⁹ ⁸⁴⁰ ⁸⁴¹ ⁸⁴² ⁸⁴³ ⁸⁴⁴ ⁸⁴⁵ ⁸⁴⁶ ⁸⁴⁷ ⁸⁴⁸ ⁸⁴⁹ ⁸⁵⁰ ⁸⁵¹ ⁸⁵² ⁸⁵³ ⁸⁵⁴ ⁸⁵⁵ ⁸⁵⁶ ⁸⁵⁷ ⁸⁵⁸ ⁸⁵⁹ ⁸⁶⁰ ⁸⁶¹ ⁸⁶² ⁸⁶³ ⁸⁶⁴ ⁸⁶⁵ ⁸⁶⁶ ⁸⁶⁷ ⁸⁶⁸ ⁸⁶⁹ ⁸⁷⁰ ⁸⁷¹ ⁸⁷² ⁸⁷³ ⁸⁷⁴ ⁸⁷⁵ ⁸⁷⁶ ⁸⁷⁷ ⁸⁷⁸ ⁸⁷⁹ ⁸⁸⁰ ⁸⁸¹ ⁸⁸² ⁸⁸³ ⁸⁸⁴ ⁸⁸⁵ ⁸⁸⁶ ⁸⁸⁷ ⁸⁸⁸ ⁸⁸⁹ ⁸⁹⁰ ⁸⁹¹ ⁸⁹² ⁸⁹³ ⁸⁹⁴ ⁸⁹⁵ ⁸⁹⁶ ⁸⁹⁷ ⁸⁹⁸ ⁸⁹⁹ ⁹⁰⁰ ⁹⁰¹ ⁹⁰² ⁹⁰³ ⁹⁰⁴ ⁹⁰⁵ ⁹⁰⁶ ⁹⁰⁷ ⁹⁰⁸ ⁹⁰⁹ ⁹¹⁰ ⁹¹¹ ⁹¹² ⁹¹³ ⁹¹⁴ ⁹¹⁵ ⁹¹⁶ ⁹¹⁷ ⁹¹⁸ ⁹¹⁹ ⁹²⁰ ⁹²¹ ⁹²² ⁹²³ ⁹²⁴ ⁹²⁵ ⁹²⁶ ⁹²⁷ ⁹²⁸ ⁹²⁹ ⁹³⁰ ⁹³¹ ⁹³² ⁹³³ ⁹³⁴ ⁹³⁵ ⁹³⁶ ⁹³⁷ ⁹³⁸ ⁹³⁹ ⁹⁴⁰ ⁹⁴¹ ⁹⁴² ⁹⁴³ ⁹⁴⁴ ⁹⁴⁵ ⁹⁴⁶ ⁹⁴⁷ ⁹⁴⁸ ⁹⁴⁹ ⁹⁵⁰ ⁹⁵¹ ⁹⁵² ⁹⁵³ ⁹⁵⁴ ⁹⁵⁵ ⁹⁵⁶ ⁹⁵⁷ ⁹⁵⁸ ⁹⁵⁹ ⁹⁶⁰ ⁹⁶¹ ⁹⁶² ⁹⁶³ ⁹⁶⁴ ⁹⁶⁵ ⁹⁶⁶ ⁹⁶⁷ ⁹⁶⁸ ⁹⁶⁹ ⁹⁷⁰ ⁹⁷¹ ⁹⁷² ⁹⁷³ ⁹⁷⁴ ⁹⁷⁵ ⁹⁷⁶ ⁹⁷⁷ ⁹⁷⁸ ⁹⁷⁹ ⁹⁸⁰ ⁹⁸¹ ⁹⁸² ⁹⁸³ ⁹⁸⁴ ⁹⁸⁵ ⁹⁸⁶ ⁹⁸⁷ ⁹⁸⁸ ⁹⁸⁹ ⁹⁹⁰ ⁹⁹¹ ⁹⁹² ⁹⁹³ ⁹⁹⁴ ⁹⁹⁵ ⁹⁹⁶ ⁹⁹⁷ ⁹⁹⁸ ⁹⁹⁹ ¹⁰⁰⁰ ¹⁰⁰¹ ¹⁰⁰² ¹⁰⁰³ ¹⁰⁰⁴ ¹⁰⁰⁵ ¹⁰⁰⁶ ¹⁰⁰⁷ ¹⁰⁰⁸ ¹⁰⁰⁹ ¹⁰¹⁰ ¹⁰¹¹ ¹⁰¹² ¹⁰¹³ ¹⁰¹⁴ ¹⁰¹⁵ ¹⁰¹⁶ ¹⁰¹⁷ ¹⁰¹⁸ ¹⁰¹⁹ ¹⁰²⁰ ¹⁰²¹ ¹⁰²² ¹⁰²³ ¹⁰²⁴ ¹⁰²⁵ ¹⁰²⁶ ¹⁰²⁷ ¹⁰²⁸ ¹⁰²⁹ ¹⁰³⁰ ¹⁰³¹ ¹⁰³² ¹⁰³³ ¹⁰³⁴ ¹⁰³⁵ ¹⁰³⁶ ¹⁰³⁷ ¹⁰³⁸ ¹⁰³⁹ ¹⁰⁴⁰ ¹⁰⁴¹ ¹⁰⁴² ¹⁰⁴³ ¹⁰⁴⁴ ¹⁰⁴⁵ ¹⁰⁴⁶ ¹⁰⁴⁷ ¹⁰⁴⁸ ¹⁰⁴⁹ ¹⁰⁵⁰ ¹⁰⁵¹ ¹⁰⁵² ¹⁰⁵³ ¹⁰⁵⁴ ¹⁰⁵⁵ ¹⁰⁵⁶ ¹⁰⁵⁷ ¹⁰⁵⁸ ¹⁰⁵⁹ ¹⁰⁶⁰ ¹⁰⁶¹ ¹⁰⁶² ¹⁰⁶³ ¹⁰⁶⁴ ¹⁰⁶⁵ ¹⁰⁶⁶ ¹⁰⁶⁷ ¹⁰⁶⁸ ¹⁰⁶⁹ ¹⁰⁷⁰ ¹⁰⁷¹ ¹⁰⁷² ¹⁰⁷³ ¹⁰⁷⁴ ¹⁰⁷⁵ ¹⁰⁷⁶ ¹⁰⁷⁷ ¹⁰⁷⁸ ¹⁰⁷⁹ ¹⁰⁸⁰ ¹⁰⁸¹ ¹⁰⁸² ¹⁰⁸³ ¹⁰⁸⁴ ¹⁰⁸⁵ ¹⁰⁸⁶ ¹⁰⁸⁷ ¹⁰⁸⁸ ¹⁰⁸⁹ ¹⁰⁹⁰ ¹⁰⁹¹ ¹⁰⁹² ¹⁰⁹³ ¹⁰⁹⁴ ¹⁰⁹⁵ ¹⁰⁹⁶ ¹⁰⁹⁷ ¹⁰⁹⁸ ¹⁰⁹⁹ ¹¹⁰⁰ ¹¹⁰¹ ¹¹⁰² ¹¹⁰³ ¹¹⁰⁴ ¹¹⁰⁵ ¹¹⁰⁶ ¹¹⁰⁷ ¹¹⁰⁸ ¹¹⁰⁹ ¹¹¹⁰ ¹¹¹¹ ¹¹¹² ¹¹¹³ ¹¹¹⁴ ¹¹¹⁵ ¹¹¹⁶ ¹¹¹⁷ ¹¹¹⁸ ¹¹¹⁹ ¹¹²⁰ ¹¹²¹ ¹¹²² ¹¹²³ ¹¹²⁴ ¹¹²⁵ ¹¹²⁶ ¹¹²⁷ ¹¹²⁸ ¹¹²⁹ ¹¹³⁰ ¹¹³¹ ¹¹³² ¹¹³³ ¹¹³⁴ ¹¹³⁵ ¹¹³⁶ ¹¹³⁷ ¹¹³⁸ ¹¹³⁹ ¹¹⁴⁰ ¹¹⁴¹ ¹¹⁴² ¹¹⁴³ ¹¹⁴⁴ ¹¹⁴⁵ ¹¹⁴⁶ ¹¹⁴⁷ ¹¹⁴⁸ ¹¹⁴⁹ ¹¹⁵⁰ ¹¹⁵¹ ¹¹⁵² ¹¹⁵³ ¹¹⁵⁴ ¹¹⁵⁵ ¹¹⁵⁶ ¹¹⁵⁷ ¹¹⁵⁸ ¹¹⁵⁹ ¹¹⁶⁰ ¹¹⁶¹ ¹¹⁶² ¹¹⁶³ ¹¹⁶⁴ ¹¹⁶⁵ ¹¹⁶⁶ ¹¹⁶⁷ ¹¹⁶⁸ ¹¹⁶⁹ ¹¹⁷⁰ ¹¹⁷¹ ¹¹⁷² ¹¹⁷³ ¹¹⁷⁴ ¹¹⁷⁵ ¹¹⁷⁶ ¹¹⁷⁷ ¹¹⁷⁸ ¹¹⁷⁹ ¹¹⁸⁰ ¹¹⁸¹ ¹¹⁸² ¹¹⁸³ ¹¹⁸⁴ ¹¹⁸⁵ ¹¹⁸⁶ ¹¹⁸⁷ ¹¹⁸⁸ ¹¹⁸⁹ ¹¹⁹⁰ ¹¹⁹¹ ¹¹⁹² ¹¹⁹³ ¹¹⁹⁴ ¹¹⁹⁵ ¹¹⁹⁶ ¹¹⁹⁷ ¹¹⁹⁸ ¹¹⁹⁹ ¹²⁰⁰ ¹²⁰¹ ¹²⁰² ¹²⁰³ ¹²⁰⁴ ¹²⁰⁵ ¹²⁰⁶ ¹²⁰⁷ ¹²⁰⁸ ¹²⁰⁹ ¹²¹⁰ ¹²¹¹ ¹²¹² ¹²¹³ ¹²¹⁴ ¹²¹⁵ ¹²¹⁶ ¹²¹⁷ ¹²¹⁸ ¹²¹⁹ ¹²²⁰ ¹²²¹ ¹²²² ¹²²³ ¹²²⁴ ¹²²⁵ ¹²²⁶ ¹²²⁷ ¹²²⁸ ¹²²⁹ ¹²³⁰ ¹²³¹ ¹²³² ¹²³³ ¹²³⁴ ¹²³⁵ ¹²³⁶ ¹²³⁷ ¹²³⁸ ¹²³⁹ ¹²⁴⁰ ¹²⁴¹ ¹²⁴² ¹²⁴³ ¹²⁴⁴ ¹²⁴⁵ ¹²⁴⁶ ¹²⁴⁷ ¹²⁴⁸ ¹²⁴⁹ ¹²⁵⁰ ¹²⁵¹ ¹²⁵² ¹²⁵³ ¹²⁵⁴ ¹²⁵⁵ ¹²⁵⁶ ¹²⁵⁷ ¹²⁵⁸ ¹²⁵⁹ ¹²⁶⁰ ¹²⁶¹ ¹²⁶² ¹²⁶³ ¹*

KW skin testing; M.tuberculosis; ss.
 XX
 XX Mycobacterium tuberculosis.
 OS
 FH Key Location/Qualifiers
 FT CDS 11..409 /tag= a
 FT /product= Antigen_TBra12
 PN MO9709429-A2.
 XX
 XX 13-MAR-1997.
 PD
 XX
 PF 30-APR-1996; 96WO-US14675.
 XX
 XX 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX
 DR WPI; 1997-192904/17.
 DR P-PSDB; AAW32354.
 XX
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX
 XX
 PS Claim 3; Page 49-50; 190pp; English.
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence encodes a specifically claimed
 CC M.tuberculosis antigen, TBra12. The immunogenic polypeptide can be
 CC used to diagnose M.tuberculosis infection by forming complexes with
 CC specific antibodies in the sample. Fragments of DNA encoding the
 CC immunogenic polypeptide can be used as diagnostic primers or probes
 CC and agents that bind to the antigen, especially monoclonal antibodies
 CC or equivalent polyclonal antibodies, are also used for diagnosis.
 CC
 XX
 PS Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;
 SQ
 Query Match 99.6%; Score 394.4; DB 18; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 acggcgcgctgcgatacttcacagctgtcccgaggtgagcagagattcgccattccgac 60
 DB 11 acggcgcgctgcgatacttcacagctgtcccgaggtgagcagagattcgccattccgac 70
 QY 61 gggcagcgatgagc 120
 DB 71 gggcagcgatgagc 130
 QY 121 atggggctaccgctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 180
 DB 131 atggggctaccgctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 190
 QY 181 gtccaaacgctggtcggagcgctccgagcgcaagctcgcgcatccacccgcgcagctg 240
 DB 191 gtccaaacgctggtcggagcgctccgagcgcaagctcgcgcatccacccgcgcagctg 250
 QY 241 ataccgcgctgacgagcctcgcgacgacgacgacgacgacgacgacgacgacgacgac 300
 DB 251 ataccgcgctgacgagcctcgcgacgacgacgacgacgacgacgacgacgacgacgac 310
 QY 301 gggcgcacatcccggtgacgtcgcgtgacgtgacgtgacgtgacgtgacgtgacgtgac 360

DB 311 gggcgcacatcccggtgacgtcgcgtgacgtgacgtgacgtgacgtgacgtgacgtgac 370
 QY 361 acagggacgtgacatggtgacgagcagcccgccgac 396
 DB 371 acagggacgtgacatggtgacgagcagcccgccgac 406
 RESULT 9
 AAV64450
 ID AAV64450 standard; DNA; 447 BP.
 XX
 AC AAV64450;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis immunogenic polypeptide TBra12 DNA.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 PD
 XX
 PF 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18293.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1998-261042/23.
 DR P-PSDB; AAW81657.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 XX
 PS Claim 3; Page 62-63; 230pp; English.
 CC This sequence encodes an immunogenic portion of a soluble Mycobacterium
 CC tuberculosis (MT) antigen which can be used in a method for inducing
 CC protective immunity against tuberculosis (TB). This sequence can be
 CC formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 CC
 XX
 PS Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;
 SQ
 Query Match 99.6%; Score 394.4; DB 19; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 acggcgcgctgcgatacttcacagctgtcccgaggtgagcagagattcgccattccgac 60
 DB 11 acggcgcgctgcgatacttcacagctgtcccgaggtgagcagagattcgccattccgac 70
 QY 61 gggcagcgatgagc 120
 DB 71 gggcagcgatgagc 130
 QY 121 atggggctaccgctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 180
 DB 131 atggggctaccgctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 190
 QY 181 gtccaaacgctggtcggagcgctccgagcgcaagctcgcgcatccacccgcgcagctg 240

```

|||||
Db 191 gtccaaagcgtgtgctcgagcgctccgcgcgaagctcgcgcatctccaccgcgacgtg 250
QY 241 ataccgcggttcgacgagcgctccgatacaactcgccacgcgcatggtgcgacgcttaac 300
Db 251 ataccgcggttcgacgagcgctccgatacaactcgccacgcgcatggtgcgacgcttaac 310
QY 301 gggcatcatcccggtgacgtcatctcgtgacctgtgcaaccaagtgcggcgacgcgct 360
Db 311 gggcatcatcccggtgacgtcatctcgtgacctgtgcaaccaagtgcggcgacgcgct 370
QY 361 acagggaacgtgacattggtccgagggaccctcgcc 396
Db 371 acagggaacgtgacattggtccgagggaccctcgcc 406

RESULT 10
AAV44342
ID AAV44342 standard; DNA; 447 BP.
XX
AC AAV44342;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen Tba12 DNA.
XX
KM Tuberculosis; infection; diagnosis; antigen; Tba12; ss.
XX
OS Mycobacterium tuberculosis strain H37Ra.
XX
PH Key Location/Qualifiers
FT CDS 11..406
FT /tag= a

W09816645-A2.
PN 23-APR-1998.
PD
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
XX
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodges MJ;
PI Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
XX
P-PSDB; AAW64294.

New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
PT
XX
PS Claim 3; Page 61; 250pp; English.
XX
CC This DNA sequence codes for Mycobacterium tuberculosis soluble
CC antigen Tba12 (see AAW64294). It was isolated from a M. tuberculosis
CC strain H37Ra expression library with rabbit anti-sera raised
CC against M. tuberculosis supernatant. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of
CC a soluble M. tuberculosis antigen, as well as DNA sequences encoding such
CC M. tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfectected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers, for the diagnosis of tuberculosis.
XX
XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other:
SQ

```

Query Match 99.6%; Score 394.4; DB 19; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 acggcgcgttcgataacttcacagctgtccaggtgtggcagagattcgcattccgattc 60
Db 11 acggcgcgttcgataacttcacagctgtccaggtgtggcagagattcgcattccgattc 70
QY 61 gggcagcgcgttcgataacttcacagctgtccaggtgtggcagagattcgcattccgattc 120
Db 71 gggcagcgcgttcgataacttcacagctgtccaggtgtggcagagattcgcattccgattc 130
QY 121 atcggttcacgccttcctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
Db 131 atcggttcacgccttcctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 190
QY 181 gtccaaagcgtgtgctcgagcgctccgcgcgaagctcgcgcatctccaccgcgacgtg 240
Db 191 gtccaaagcgtgtgctcgagcgctccgcgcgaagctcgcgcatctccaccgcgacgtg 250
QY 241 ataccgcggttcgacgagcgctccgatacaactcgccacgcgcatggtgcgacgcttaac 300
Db 251 ataccgcggttcgacgagcgctccgatacaactcgccacgcgcatggtgcgacgcttaac 310
QY 301 gggcatcatcccggtgacgtcatctcgtgacctgtgcaaccaagtgcggcgacgcgct 360
Db 311 gggcatcatcccggtgacgtcatctcgtgacctgtgcaaccaagtgcggcgacgcgct 370
QY 361 acagggaacgtgacattggtccgagggaccctcgcc 396
Db 371 acagggaacgtgacattggtccgagggaccctcgcc 406

RESULT 11
AAZ19252
ID AAZ19252 standard; DNA; 447 BP.
XX
AC AAZ19252;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen Tba12 DNA sequence.
XX
KM Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KM immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN W09942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
XX
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodges MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.
XX
P-PSDB; AAY39096.

New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
PT
XX
PS Claim 3; Page 73; 299pp; English.
XX
CC The present invention describes polypeptides comprising an immunogenic
CC

```

CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to
CC AA219225 are used in the exemplification of the present invention.
XX
SQ Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;

Query Match 99.6%; Score 394.4; DB 20; Length 447;
Best Local Similarity 99.7%; Pred. No. 4.7e-85;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggccggtccgataactccagctgtcccaagggtggcaggagatccgcatccgac 60
DB 11 acggccggtccgataactccagctgtcccaagggtggcaggagatccgcatccgac 70
QY 61 gggcagcgatggcgatccgcggccagatccggtgggggggtccaccacggtcat 120
DB 71 gggcagcgatggcgatccgcggccagatccggtgggggggtccaccacggtcat 130
QY 121 atcgagcctacgctctctcgtggtgtgttcgaaacaaacggcagcgacaga 180
DB 131 atcgagcctacgctctctcgtggtgtgttcgaaacaaacggcagcgacaga 190
QY 181 gtcacagcggtgctggcgagcgctccgcggcgaagltcgcgcatccacggcgacgtg 240
DB 191 gtcacagcggtgctggcgagcgctccgcggcgaagltcgcgcatccacggcgacgtg 250
QY 241 atcacggcggtgagcgagcgctccgataactccgacggcgatggcgagcggttac 300
DB 251 atcacggcggtgagcgagcgctccgataactccgacggcgatggcgagcggttac 310
QY 301 gggcagcctacgggtgagcgtcgtgacgtggcgaacaaacggtggcgacggt 360
DB 311 gggcagcctacgggtgagcgtcgtgacgtggcgaacaaacggtggcgacggt 370
QY 361 acagggaacgtgacatggcgagggagaccccgccg 396
DB 371 acagggaacgtgacatggcgagggagaccccgccg 406

RESULT 12

AA219040
ID AA219040 standard; DNA; 447 BP.

AC AA219040;
XX
DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen DNA encoding TbrA12.

KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity; ss.

OS Mycobacterium tuberculosis.

PN WO9942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

PA (COR-) CORIXA CORP.
XX

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R,
PI Lodes MW, Reed SG, Skelky YAW, Twardzik DR, Vadvick TS;
XX
XX WP1; 1999-527416/44.
DR P-SDB; AAY38959.

PT New polypeptide comprising antigenic portions of M. tuberculosis
XX
XX
PS Claim 3; Page 103; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.

SQ Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;

Query Match 99.6%; Score 394.4; DB 20; Length 447;
Best Local Similarity 99.7%; Pred. No. 4.7e-85;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggccggtccgataactccagctgtcccaagggtggcaggagatccgcatccgac 60
DB 11 acggccggtccgataactccagctgtcccaagggtggcaggagatccgcatccgac 70
QY 61 gggcagcgatggcgatccgcggccagatccggtgggggggtccaccacggtcat 120
DB 71 gggcagcgatggcgatccgcggccagatccggtgggggggtccaccacggtcat 130
QY 121 atcgagcctacgctctctcgtggtgtgttcgaaacaaacggcagcgacaga 180
DB 131 atcgagcctacgctctctcgtggtgtgttcgaaacaaacggcagcgacaga 190
QY 181 gtcacagcggtgctggcgagcgctccgcggcgaagltcgcgcatccacggcgacgtg 240
DB 191 gtcacagcggtgctggcgagcgctccgcggcgaagltcgcgcatccacggcgacgtg 250
QY 241 atcacggcggtgagcgagcgctccgataactccgacggcgatggcgagcggttac 300
DB 251 atcacggcggtgagcgagcgctccgataactccgacggcgatggcgagcggttac 310
QY 301 gggcagcctacgggtgagcgtcgtgacgtggcgaacaaacggtggcgacggt 360
DB 311 gggcagcctacgggtgagcgtcgtgacgtggcgaacaaacggtggcgacggt 370
QY 361 acagggaacgtgacatggcgagggagaccccgccg 396
DB 371 acagggaacgtgacatggcgagggagaccccgccg 406

RESULT 13

AA203780
ID AA203780 standard; DNA; 447 BP.

AC AA203780;

DT 29-AUG-2001 (first entry)

DE M. tuberculosis DNA encoding a partial antigen TbrA12.

KW TbrA12; antigen; vaccine; tuberculosis; AIDS;
KW acquired immunodeficiency disease; ss.

OS Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT 11. 409 /tag= a "TbrA12"

FT /product= "TbrA12"

```

FT FT /partial
XX XX /note="No start codon"
PN WO200124820-A1.
PD 12-APR-2001.
PE 10-OCT-2000; 2000WO-US28095.
PR 07-OCT-1999; 99US-0158338.
XX 07-OCT-1999; 99US-0158425.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX WPI: 2001-290576/30.
XX P-PSDB: AAU01889.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
XX proteins comprising combinations of heterologous antigens
XX
XX Example 2; Page 162; 168pp; English.
XX
XX The sequence encodes Mycobacterium tuberculosis TbrA12,
XX an M. tuberculosis antigen. Compositions comprising at least
XX 2 heterologous antigens, as a fusion protein, and vectors expressing the
XX fusion proteins are used as vaccines to prophylactically immunise
XX mammals (especially humans) against infection by Mycobacteria. The
XX compositions contain at least 2 heterologous antigens that increase the
XX serological sensitivity of individuals infected with tuberculosis, a
XX disease frequently affecting patients with acquired immunodeficiency
XX disease, AIDS.
XX
XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;
SQ

```

```

Query Match 99.6%; Score 394.4; DB 22; Length 447;
Best Local Similarity 99.7%; Pred. No. 4,7e-85;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 acggcggtcgcgaatactccagctgtccagggtggcgaggtccatccgcatc 60
DB 11 acggcggtcgcgaatactccagctgtccagggtggcgaggtccatccgcatc 70
OY 61 ggccagcgtatggtcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
DB 71 ggccagcgtatggtcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 130
OY 121 atcggcgtacgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 180
DB 131 atcggcgtacgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 190
OY 181 gtccaaacgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 240
DB 191 gtccaaacgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 250
OY 241 atcaccgggtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
DB 251 atcaccgggtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 310
OY 301 ggccatcctccgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 360
DB 311 ggccatcctccgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 370
OY 361 acagggaacgtgacatgtggtcggagggagcccgccgccc 396
DB 371 acagggaacgtgacatgtggtcggagggagcccgccgccc 406

```

```

RESULT 14
AAS03793
ID AAS03793 standard; DNA; 1629 BP.

```

```

XX XX
XX AAS03793;
XX
XX 29-AUG-2001 (first entry)
XX
XX M. tuberculosis DNA encoding antigen HTCC#1 fusion protein #4.
XX
XX TbrA12-HTCC#1; antigen; vaccine; tuberculosis;
XX AIDS; acquired immunodeficiency disease; His tag; ds.
XX
XX Mycobacterium tuberculosis.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..1629
XX
XX /tag= a "TbrA12-HTCC#1"
XX /product= (pos:1621..1623,aa:Xaa)
XX /transl_except= (pos:1621..1623,aa:Xaa)
XX /note= "Xaa" in frame STOP codon"
XX
XX /partial
XX /note= "No start or stop codon. Although the sequence
XX does contain an in frame stop codon, 2 further amino
XX acids are shown in figure 8 as being encoded by the
XX present sequence, without a further stop codon"
XX 25..426
XX /tag= b
XX /note= "Region derived from TbrA12"
XX
XX /tag= c
XX /note= "Region derived from Thrombin"
XX
XX /tag= d
XX /note= "Region derived from HTCC#1"
XX
XX WO200124820-A1.
XX
XX 12-APR-2001.
XX
XX 10-OCT-2000; 2000WO-US28095.
XX
XX 07-OCT-1999; 99US-0158338.
XX
XX 07-OCT-1999; 99US-0158425.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX WPI: 2001-290576/30.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
XX proteins comprising combinations of heterologous antigens
XX
XX Example 2; Fig 8; 168pp; English.
XX
XX The sequence encodes Mycobacterium tuberculosis fusion protein,
XX TbrA12-HTCC#1 and includes a His tag at the N-terminus to aid
XX purification. Compositions comprising at least 2 heterologous antigens,
XX as a fusion protein, and vectors expressing the fusion proteins are
XX used as vaccines to prophylactically immunise mammals (especially
XX humans) against infection by Mycobacteria. The compositions
XX contain at least 2 heterologous antigens that increase the serological
XX sensitivity of individuals infected with tuberculosis, a disease
XX frequently affecting patients with acquired immunodeficiency disease,
XX AIDS.
XX
XX Sequence 1629 BP; 298 A; 500 C; 535 G; 296 T; 0 other;
SQ

```

```

Query Match 99.6%; Score 394.4; DB 22; Length 1629;
Best Local Similarity 99.7%; Pred. No. 5.2e-85;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 acggcggtcgcgaatactccagctgtccagggtggcgaggtccatccgcatc 60

```

```

Db      25  acgcccgcgtccgataactccagctgtccccaggtgtgaggaattccgcatccgcatc 84
QY      61  gggcagcgatggtcgatccgagccagatccgagtcggtggggggtccaccccgcttat 120
Db      85  gggcagcgatggtcgatccgagccagatccgagtcggtggggggtccaccccgcttat 144
QY     121  atcgagcctacgcgcttccctcggtgtgtgttcgcagacaaacgagcagcagcga 180
Db     145  atcgagcctacgcgcttccctcggtgtgtgttcgcagacaaacgagcagcagcga 204
QY     181  gtccacgcgtgtgtcgagcgctccgagcgagtcctcgatccatccacccgagcgtg 240
Db     205  gtccacgcgtgtgtcgagcgctccgagcgagtcctcgatccatccacccgagcgtg 264
QY     241  atcacgcggtgtcgagcgctccgagtcctcgatccatccacccgagcgtgagcgttaac 300
Db     265  atcacgcggtgtcgagcgctccgagtcctcgatccatccacccgagcgtgagcgttaac 324
QY     301  gggcatalatcccggtgagcgtatcctcggtgacctggaacaaacgagtcgagcgagcgt 360
Db     325  gggcatalatcccggtgagcgtatcctcggtgacctggaacaaacgagtcgagcgagcgt 384
QY     361  acagggagcgtgacatgtgagcgagagaccccgagcc 396
Db     385  acagggagcgtgacatgtgagcgagagaccccgagcc 420

RESULT  15
AAT91477
ID      AAT91477 standard; DNA; 1872 BP.
AC      AAT91477;
XX
DT      07-JAN-1998 (first entry)
XX
DE      Mycobacterium tuberculosis antigen TBRA35 encoding DNA.
XX
KW      Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX      skin testing; M.tuberculosis; ss.
XX
OS      Mycobacterium tuberculosis.
XX
FH      Key      Location/Qualifiers
FT      CDS      89..1156
FT              /tag=a
FT              /product= Antigen_TBRA35
FT
XX
PN      W09709428-A2.
XX
PD      13-MAR-1997.
XX
PE      30-AUG-1996; 96WO-US14674.
XX
PR      12-JUL-1996; 96US-0680574.
PR      01-SEP-1995; 95US-0523436.
PR      22-SEP-1995; 95US-0533634.
PR      22-MAR-1996; 96US-0620874.
PR      05-JUN-1996; 96US-0659683.
XX
PA      (CORI-) CORIXA CORP.
XX
PI      Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAM;
PI      Twardzik DR, Vedvyck TH;
XX
DR      WPI; 1997-192903/17.
DR      P-PSDB; AAW32435.
XX
PT      New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT      useful in vaccines for prevention or treatment of tuberculosis, also
PT      for diagnosis
XX
PS      Claim 3; Page 63-65; 168pp; English.

```

```

XX      A new immunogenic polypeptide has been developed comprising an
CC      immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC      its variant differing only in conservative substitutions and/or
CC      modifications). The present sequence encodes a specifically claimed
CC      M.tuberculosis antigen, TBRA35. The immunogenic protein, and fusion
CC      proteins containing one or more of the proteins or one of the proteins
CC      plus ESAT-6, are useful in vaccines, preferably when formulated with a
CC      non-specific adjuvant, to induce an immune response against
CC      M.tuberculosis (for treatment or prevention).
XX
SQ      Sequence 1872 BP; 318 A; 616 C; 604 G; 332 T; 2 other:

Query Match          98.8%; Score 391.2; DB 18; Length 1872;
Best Local Similarity 99.2%; Pred. No. 3,1e-84;
Matches 393; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  acgcccgcgtccgataactccagctgtccccaggtgtgaggaattccgcatccgcatc 60
Db      758  acgcccgcgtccgataactccagctgtccccaggtgtgaggaattccgcatccgcatc 817
QY      61  gggcagcgatggtcgatccgagccagatccgagtcggtggggggtccaccccgcttat 120
Db      818  gggcagcgatggtcgatccgagccagatccgagtcggtggggggtccaccccgcttat 877
QY     121  atcgagcctacgcgcttccctcggtgtgtgttcgcagacaaacgagcagcagcga 180
Db      878  atcgagcctacgcgcttccctcggtgtgtgttcgcagacaaacgagcagcagcga 937
QY     181  gtccacgcgtgtgtcgagcgctccgagcgagtcctcgatccatccacccgagcgtg 240
Db      938  gtccacgcgtgtgtcgagcgctccgagcgagtcctcgatccatccacccgagcgtg 997
QY     241  atcacgcggtgtcgagcgctccgagtcctcgatccatccacccgagcgtgagcgttaac 300
Db      998  atcacgcggtgtcgagcgctccgagtcctcgatccatccgacacgagtcgagcagcgttaac 1057
QY     301  gggcatalatcccggtgagcgtatcctcggtgacctggaacaaacgagtcgagcgagcgt 360
Db     1058  gggcatalatcccggtgagcgtatcctcggtgacctggaacaaacgagtcgagcgagcgt 1117
QY     361  acagggagcgtgacatgtgagcgagagaccccgagcc 396
Db     1118  acagggagcgtgacatgtgagcgagagaccccgagcc 1153

```

Search completed: August 6, 2002, 11:39:56
Job time: 6866 sec

DR WPI, 1999-601610/51.
DR N-PSDB; AA220206.

PT New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -

PS Claim 1; Fig 13A-B; 83pp; English.

CC This sequence represents a recombinant Mycobacterium tuberculosis
CC bi-antigen fusion protein, termed MtB24, composed of the antigens
CC R12 and DPD. The fusion protein is expressed in host cells
CC using a vector carrying a polynucleotide (see AA220206) comprising
CC the coding sequences for the 2 antigens. The invention provides
CC fusion proteins (see AA22059-71) containing at least 2 M.
CC tuberculosis antigens. The new fusion proteins and polynucleotides
CC encoding them are useful as vaccines for preventing tuberculosis
CC (claimed), for diagnosis (via in vitro assays or intradermal skin
CC tests for detection of anti-M. tuberculosis antibodies), monitoring
CC of disease progression, and treatment of tuberculosis. They are
CC more effective immunogens than mixtures of the individual protein
CC components.

XX Sequence 231 AA;

Query Match 100.0%; Score 675; DB 20; Length 231;
Best Local Similarity 100.0%; Pred. No. 4.8e-59;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGCGFAIPIGOMAIAGQIRSGGSPYHIGPTAFGLGVVNNNGNGAR 60
Db 9 taasdnfqlsggggfgfapigamalaqfrrsggsptvhiptatlglyvndngngar 68

QY 61 VORVVSAPASISGISTGVITAVDGPINSATAMADALNGHHPGVISVTWQTSGGTR 120
Db 69 vqrvvsgapasaigistgdivitavdgapinsatamadalinghpgdivisvtwqlksygr 128

QY 121 TGNVTLAEGPPA 132
Db 129 tgnvltlaegppa 140

RESULT 2

AAV05000 standard; Protein; 355 AA.

AC AAV05000;

DT 06-JUL-1999 (first entry)

DE Mycobacterium species protein sequence 50D.

KM Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KM hybridisation; detection; vaccine; immunisation; infection.

OS Mycobacterium sp.

PN WO9090186-A2.

PD 25-FEB-1999.

PF 14-AUG-1998; 98WO-FR01813.

PR 11-SEP-1997; 97FR-0011325.

PR 14-AUG-1997; 97FR-0010404.

PA (INSP) INSR PASTEUR.

PI Gicquel B, Lhm EM, Pelicic V, Portnoi D, Goguel de la Salmoniere Y;

DR WPI, 1999-181045/15.
DR N-PSDB; AA234251.

XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression

PS Claim 32; Fig 50D; 309pp; French.

CC Sequences AAV04742-Y05000 and AAV07201-Y07204 represent secreted
CC proteins from various Mycobacterium species microorganisms. The
CC encoding nucleotide sequences can be used as primers and probes for
CC methods for detecting and identifying mycobacteria, especially belonging
CC to the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.

XX Sequence 355 AA;

Query Match 100.0%; Score 675; DB 20; Length 355;
Best Local Similarity 100.0%; Pred. No. 8.3e-59;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGCGFAIPIGOMAIAGQIRSGGSPYHIGPTAFGLGVVNNNGNGAR 60
Db 224 taasdnfqlsggggfgfapigamalaqfrrsggsptvhiptatlglyvndngngar 283

QY 61 VORVVSAPASISGISTGVITAVDGPINSATAMADALNGHHPGVISVTWQTSGGTR 120
Db 284 vqrvvsgapasaigistgdivitavdgapinsatamadalinghpgdivisvtwqlksygr 343

QY 121 TGNVTLAEGPPA 132
Db 344 tgnvltlaegppa 355

RESULT 3

AAG81110 standard; Protein; 355 AA.

AC AAG81110;

DT 04-SEP-2001 (first entry)

DE Mycobacterium tuberculosis potential drug target protein SEQ ID 161.

KM Drug target; growth; organism viability; characterisation.

OS Mycobacterium tuberculosis.

PN WO200135317-A1.

PD 17-MAY-2001.

PF 13-NOV-2000; 2000WO-US31152.

PR 12-NOV-1999; 99US-0165086.

PR 12-NOV-1999; 99US-0165124.

PR 01-FEB-2000; 2000US-0179531.

PA (REGC) UNIV CALIFORNIA.

PI Eisenberg D, Rotstein SH, Marcotte EM;

DR WPI, 2001-329193/34.

DR N-PSDB; AAH51961.

PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences -

PS Disclosure; Page 157; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or
CC

CC antigens, as a fusion protein, and vectors expressing the fusion
 CC proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS.
 XX
 SQ Sequence 543 AA;

Query Match 100.0%; Score 675; DB 22; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.4e-58;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGAR 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9 taasdnfqlsggggffalpgamalaagqlrsggspvthigpafglgvndngngar 68
 QY 61 VQRYVGSAPAAISIGTGDVITAVDGPINSATAMADALNGHHPGDVTSVTWTKSGGTR 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 69 vqrvvgsapaaisigstgdlvavdgapinsatamadalinghpgdvlsvtwtksggtr 128
 QY 121 TGNVTLAEGPPA 132
 ||||||||||||
 Db 129 tgnvtlaegppa 140

RESULT 6

AAV32059 6
 ID AAV32059 standard; Protein; 729 AA.

AC AAV32059;

DT 17-JAN-2000 (first entry)

XX Mycobacterium tuberculosis antigen fusion protein MtB32A.

KW Tuberculosis; antigen; fusion protein; MtB32A; Ra12; Tbh9; Ra35;

KM diagnosis; therapy; vaccine; immunogen.

XX Mycobacterium tuberculosis.

FT Key Location/Qualifiers

FT Peptide 1..7 /note= "Met/His tag"

FT Protein 8..139 /note= "Ra12"

FT Protein 142..532 /note= "Tbh9"

FT Protein 535..729 /note= "Ra35"

FT Protein /note= "Ra35"

XX WO951748-A2.

XX 14-OCT-1999.

XX 07-APR-1999; 99WO-US07717.

XX 07-APR-1998; 98US-0056556.

XX 30-DEC-1998; 98US-0223040.

XX (CORI-) CORIXA CORP.

XX Skelky YAW, Alderson M, Campos-Neto A;

XX WPI; 1999-601610/51.

XX N-PSDB; AAZ20194.

XX New fusion proteins useful for diagnosis, prevention and treatment of

XX tuberculosis -

XX Claims 1; Fig 1A-B; 83pp; English.

XX This sequence represents a recombinant Mycobacterium tuberculosis
 CC tri-antigen fusion protein, termed MtB32A, composed of the antigens
 CC Ra12, Tbh9 and Ra35. The fusion protein is expressed in host cells
 CC using a vector carrying a polynucleotide (see AAZ20194) comprising
 CC the 3 coding sequences for the antigens. The invention provides
 CC fusion proteins (see AAV32059-71) containing at least 2 M.
 CC tuberculosis antigens. The new fusion proteins and polynucleotides
 CC encoding them are useful as vaccines for preventing tuberculosis
 CC (claimed), for diagnosis (via in vitro assays or intradermal skin
 CC tests for detection of anti-M. tuberculosis antibodies), monitoring
 CC of disease progression, and treatment of tuberculosis. They are
 CC more effective immunogens than mixtures of the individual protein
 CC components.
 XX
 SQ Sequence 729 AA;

Query Match 100.0%; Score 675; DB 20; Length 729;
 Best Local Similarity 100.0%; Pred. No. 2.1e-58;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGAR 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 8 taasdnfqlsggggffalpgamalaagqlrsggspvthigpafglgvndngngar 67
 QY 61 VQRYVGSAPAAISIGTGDVITAVDGPINSATAMADALNGHHPGDVTSVTWTKSGGTR 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 68 vqrvvgsapaaisigstgdlvavdgapinsatamadalinghpgdvlsvtwtksggtr 127
 QY 121 TGNVTLAEGPPA 132
 ||||||||||||
 Db 128 tgnvtlaegppa 139

RESULT 7

AAW32422 7
 ID AAW32422 standard; Protein; 132 AA.

AC AAW32422;

DT 08-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen TbhA12.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KM skin testing; M. tuberculosis.

XX Mycobacterium tuberculosis.

XX WO9709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.

XX 12-JUL-1996; 96US-0680574.

XX 01-SEP-1995; 95US-0523436.

XX 22-SEP-1995; 95US-053634.

XX 22-MAR-1996; 96US-0620874.

XX 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skelky YAW;

XX Twardzik DR, Veddick TH.

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91466.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are

XX useful in vaccines for prevention or treatment of tuberculosis, also

XX for diagnosis

antibodies, or oligonucleotide probes and primers, for the

XX
XX

U N C E R T A I N

SQ Sequence 132 AA;

Query Match 99.3%; Score 670; DB 20; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPFVHIGPTAFGLGVVDNNGNGAR 60
 |||||||
 Db 1 taasdnfqlsggggqfaipigqamalaqqlrsggspvthigpqlfqlgvvdnngngar 60
 |||||||

OY 61 VORVGSAPASISIGISTGDVITAVDGAIPINSATAMADALNGHHHPGDVISTWOTKSGCTR 120
 |||||||
 Db 61 vqrvvgsapaasligistgdtavdgapinsatamadalinghpgdvistvwtksgtr 120
 |||||||

OY 121 TGNVTLAEGPPA 132
 |||||||
 Db 121 tgnvllaegppa 132

RESULT 12
 AAY38959
 ID AAY38959 standard; Protein; 132 AA.
 XX
 AC AAY38959;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen protein TbrA12.
 XX
 KM Antigen; diagnosis; detection; infection; antibody; immunisation;
 KM vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PE 17-FEB-1999; 99MO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AA219040.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX
 PS Example 3; Page 138; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 CC
 XX
 XX Sequence 132 AA;
 SQ

Query Match 99.3%; Score 670; DB 20; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPFVHIGPTAFGLGVVDNNGNGAR 60
 |||||||
 Db 1 taasdnfqlsggggqfaipigqamalaqqlrsggspvthigpqlfqlgvvdnngngar 60
 |||||||

OY 61 VORVGSAPASISIGISTGDVITAVDGAIPINSATAMADALNGHHHPGDVISTWOTKSGCTR 120
 |||||||
 Db 61 vqrvvgsapaasligistgdtavdgapinsatamadalinghpgdvistvwtksgtr 120
 |||||||

OY 121 TGNVTLAEGPPA 132
 |||||||
 Db 121 tgnvllaegppa 132

RESULT 13
 AAU69898
 ID AAU69898 standard; Protein; 132 AA.
 XX
 AC AAU69898;
 XX
 DT 30-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Ra12.
 XX
 KM Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;
 KM fusion protein; Ra12 antigen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W0200173032-A2.
 XX
 PD 04-OCT-2001.
 XX
 PE 27-MAR-2001; 2001WO-US09919.
 PF
 XX 27-MAR-2000; 2000US-0536857.
 PR 09-MAY-2000; 2000US-0568100.
 PR 12-MAY-2000; 2000US-0570737.
 PR 13-JUN-2000; 2000US-0593793.
 PR 27-JUN-2000; 2000US-0605783.
 PR 10-AUG-2000; 2000US-0636215.
 PR 29-AUG-2000; 2000US-0651236.
 PR 06-SEP-2000; 2000US-0657279.
 PR 02-OCT-2000; 2000US-0679426.
 PR 10-OCT-2000; 2000US-0685166.
 XX
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Reltter MM, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
 XX
 DR WPI; 2001-639232/73.
 DR
 XX
 PT New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer -
 XX
 PS Example 17; Page 531-532; 579pp; English.
 XX
 CC The invention relates to isolated prostate-specific
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumour protein, and for inhibiting the development
 CC of cancer especially prostate cancer. Compositions comprising the
 CC polynucleotide and/or polypeptide are useful for stimulating an immune
 CC response, and for treating cancer. The oligonucleotide is useful for
 CC detecting cancer. The present sequence is M. tuberculosis Ra12
 CC antigen (or fragment) used to make a fusion protein comprising a
 CC prostate specific polypeptide of the invention.
 CC
 XX
 XX Sequence 132 AA;
 SQ

Query Match 99.3%; Score 670; DB 22; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGEFAIPICGAMATAIGQIRSGGSPVHIGTAFGLGVVNDNGNGAR 60
 |||||||
 Db 1 taasdnfqlsgggqgfaiipigamataigqirsggspvhlipglfqlgvvndngngar 60
 QY 61 VORVVGSAFASISIGTSDVITAVDGAPINSATAMADALNGHHPGDVISTVOTKSGGTR 120
 |||||||
 Db 61 vgrvvgssapaaslgistgdviltavdgapinsatamadalnghpgdvisvnmqlksqgr 120
 QY 121 TGNVTLAEGPPA 132
 |||||||
 Db 121 tgnvtlaegppa 132

RESULT 14

AAU69906 standard; protein; 132 AA.

AAU69906;

30-JAN-2002 (first entry)

Mycobacterium tuberculosis antigen Ra12.

Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;

fusion protein; Ra12 antigen.

Mycobacterium tuberculosis.

WO200173032-A2.

04-OCT-2001.

27-MAR-2001; 2001WO-US09919.

27-MAR-2000; 2000US-0536857.
 09-MAY-2000; 2000US-0568100.
 12-MAY-2000; 2000US-0570737.
 13-JUN-2000; 2000US-0593793.
 27-JUN-2000; 2000US-0605783.
 10-AUG-2000; 2000US-0636215.
 23-AUG-2000; 2000US-0651236.
 06-SEP-2000; 2000US-0657279.
 02-OCT-2000; 2000US-0679426.
 10-OCT-2000; 2000US-0685166.

(CORI-) CORIXA CORP.

Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;

WPI; 2001-639232/73.

New human prostate-specific polypeptides and polynucleotides useful for
 the diagnosis and treatment of cancer, especially prostate cancer -
 Example 17; Page 541-542; 579pp; English.

The invention relates to isolated prostate-specific
 polynucleotides, polypeptides, fusion proteins of the polypeptides,
 antibodies raised against the polypeptides (or antigenic epitopes
 derived from them) and antigen-presenting cells expressing the
 polypeptides. The antibodies are useful for detecting the presence of
 cancer, especially prostate cancer. The polypeptides, polynucleotides and
 the antigen-presenting cells are useful for stimulating and/or expanding
 T cells specific for a tumour protein, and for inhibiting the development
 of cancer, especially prostate cancer. Compositions comprising the
 polynucleotide and/or polypeptide are useful for stimulating an immune
 response, and for treating cancer. The oligonucleotide is useful for

detecting cancer. The present sequence is M. tuberculosis Ra12
 antigen (or fragment) used to make a fusion protein comprising a
 prostate specific polypeptide of the invention.

Sequence 132 AA;

Query Match 99.3%; Score 670; DB 22; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGEFAIPICGAMATAIGQIRSGGSPVHIGTAFGLGVVNDNGNGAR 60
 |||||||
 Db 1 taasdnfqlsgggqgfaiipigamataigqirsggspvhlipglfqlgvvndngngar 60
 QY 61 VORVVGSAFASISIGTSDVITAVDGAPINSATAMADALNGHHPGDVISTVOTKSGGTR 120
 |||||||
 Db 61 vgrvvgssapaaslgistgdviltavdgapinsatamadalnghpgdvisvnmqlksqgr 120
 QY 121 TGNVTLAEGPPA 132
 |||||||
 Db 121 tgnvtlaegppa 132

RESULT 15

AAU01253 standard; Protein; 132 AA.

AAU01253;

04-OCT-2001 (first entry)

Mycobacterium tuberculosis antigen Ra12 amino acids.

Human; prostate cancer; prostate-specific; diagnosis; vaccine;
 cytostatic; gene therapy; metastasis.

Mycobacterium tuberculosis.

WO200151633-A2.

19-JUL-2001.
 16-JAN-2001; 2001WO-US01574.
 14-JAN-2000; 2000US-0483672.

(CORI-) CORIXA CORP.

Xu J, Dillon DC, Mitcham JL, Harlocker SL, Reed SG;
 Kalos MD, Fanger GR, Retter MW, Stolk JA, Skeiky YAW;
 Wang A, Meagher MJ;
 WPI; 2001-425873/45.

New polynucleotide encoding a prostate-specific protein, for
 diagnosing, monitoring and treating prostate cancer in a patient and
 for use in vaccines -
 Example 17; Page 492; 543pp; English.

The present invention describes polynucleotide sequences (I) which encode
 prostate-specific proteins (II). (I) and (II) have cytostatic activity,
 and can be used in vaccine production and gene therapy. (I), (II),
 antibodies to (II), fusion proteins comprising (II), and isolated
 T cells prepared using (I) or (II) are used to treat cancer in a patient.
 (I) and the antibodies are also used in the detection of cancer in a
 patient. The cancer that is diagnosed or treated is particularly
 prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
 (I) can be used for monitoring the progression of cancer in a patient.
 (I) and (II) can also be used to improve diagnostic and therapeutic
 methods for prostate cancer. They can indicate the level of metastasis
 as well as the prostate volume. AAH93357 to AAH93944 and AAU01115 to

CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.

CC exemplification of the present invention.

XX

SQ Sequence 132 AA;

| Query Match | Score | DB | Length |
|-------------|-------|----|--------|
| 99.38 | 670 | 22 | 132 |

Best Local Similarity 99.28; Pred. No. 7.4e-59;

Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFQLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGLGVVDNNGNGAR 60

Db 1 taasdnfqlsqggqfaipigqamalaqgirsqggsptvhigptaflglgvndnngngar 60

0Y 61 VQRVVGSA PASLGISTGDVITAVDGA PINSATAMADALNGHPGDVISVTWQTKSGGR 120

Db 61 vqrivvsapaasigistgdtavdgapinsaramadalngihpgdtvsnwqtksgytr 120

| | | | |
|----|-----|--------------|-----|
| QY | 121 | TGNVTLAEGPPA | 132 |
|----|-----|--------------|-----|

Db 121 tgnvtlaegppa 132

Search completed: August 6, 2002, 08:15:50
Job time: 450 sec

Job time: 450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:12:40 ; Search time 19.71 Seconds
(without alignments) 643.521 Million cell updates/sec

Title: US-09-684-215A-4

Perfect score: 675
Sequence: 1 TAAADNFQLSGGGGAFAIP1.....QTKSGTRKGNVTLAEGRPA 132

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match Length | DB ID | Description |
|------------|-------|-------|--------------|-------|-------------|
| 1 | 675 | 100.0 | 355 | 2 | F70983 |
| 2 | 474.5 | 70.3 | 361 | 2 | S47170 |
| 3 | 459.5 | 68.1 | 354 | 2 | A87242 |
| 4 | 199.5 | 29.6 | 464 | 2 | C70821 |
| 5 | 188.5 | 27.9 | 452 | 2 | H86930 |
| 6 | 188.5 | 27.9 | 452 | 2 | T45448 |
| 7 | 139 | 20.6 | 394 | 2 | S74643 |
| 8 | 139 | 20.6 | 407 | 2 | AG2150 |
| 9 | 138.5 | 20.5 | 362 | 2 | T35287 |
| 10 | 133.5 | 19.8 | 408 | 2 | H86881 |
| 11 | 131 | 19.4 | 441 | 2 | E53357 |
| 12 | 128.5 | 19.0 | 514 | 2 | A82581 |
| 13 | 122.5 | 18.1 | 474 | 2 | F83550 |
| 14 | 119.5 | 17.7 | 203 | 2 | T35866 |
| 15 | 117.5 | 17.4 | 530 | 2 | F87590 |
| 16 | 113 | 16.7 | 393 | 2 | E95261 |
| 17 | 113 | 16.7 | 397 | 2 | B98127 |
| 18 | 113 | 16.7 | 513 | 2 | I40060 |
| 19 | 113 | 16.7 | 513 | 2 | AD3418 |
| 20 | 113 | 16.7 | 523 | 2 | A97479 |
| 21 | 113 | 16.7 | 523 | 2 | A12696 |
| 22 | 112 | 16.6 | 429 | 2 | AD1894 |
| 23 | 111 | 16.4 | 455 | 2 | C91142 |
| 24 | 111 | 16.4 | 455 | 2 | F85987 |
| 25 | 111 | 16.4 | 499 | 2 | B81914 |
| 26 | 109 | 16.1 | 455 | 2 | A80909 |
| 27 | 108 | 16.0 | 348 | 2 | H96956 |
| 28 | 108 | 16.0 | 472 | 2 | C87408 |
| 29 | 107.5 | 15.9 | 452 | 2 | S77538 |

| | | | | | | |
|----|-------|------|-----|---|--------|---------------------|
| 30 | 106.5 | 15.8 | 355 | 1 | JC6052 | trypsin-like prote |
| 31 | 106.5 | 15.8 | 355 | 2 | D91142 | proteinase [import |
| 32 | 106.5 | 15.8 | 355 | 2 | G85987 | proteinase [import |
| 33 | 106.5 | 15.7 | 455 | 2 | JC6051 | trypsin-like prote |
| 34 | 105 | 15.6 | 475 | 1 | S15337 | heat shock protein |
| 35 | 104.5 | 15.5 | 362 | 2 | A10433 | proteinase (EC 3.4 |
| 36 | 102.5 | 15.2 | 474 | 2 | S45229 | proteinase DO (EC |
| 37 | 102.5 | 15.2 | 474 | 2 | E83500 | proteinase DO (EC |
| 38 | 102.5 | 15.2 | 474 | 2 | E90649 | proteinase DO (EC |
| 39 | 102 | 15.1 | 475 | 2 | AC0528 | proteinase DO precu |
| 40 | 100.5 | 14.9 | 453 | 1 | B70426 | periplasmic serine |
| 41 | 100.5 | 14.9 | 459 | 2 | F72359 | periplasmic serine |
| 42 | 100 | 14.8 | 398 | 2 | B71284 | probable periplasm |
| 43 | 100 | 14.8 | 652 | 2 | E97857 | cell surface antag |
| 44 | 99.5 | 14.7 | 224 | 2 | F70048 | serine proteinase |
| 45 | 98.5 | 14.6 | 481 | 2 | D82826 | heat shock protein |

ALIGNMENTS

RESULT 1
F70983
probable serine proteinase pepa - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70983
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A: Reference number: A70500; M0ID: 98295987
A: Accession: F70983
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1355 <COL>
A: Cross-references: GB:296071, GB:AL123456, NID:q3242254, P1DN:CA809453.1, P1D:q21819
A: Experimental source: Strain H37RV
A: Genetics:
A: Gene: pepa
A: Superfamily: Escherichia coli trypsin-like proteinase

Query Match 100.0%; Score 675; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 9, 2e-49;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAADNFQLSGGGGAFAIP1GQAMATAGQIRSGGSPVTHIGTFAELGLGVVDNNGNGAR 60
Db 224 TAAADNFQLSGGGGAFAIP1GQAMATAGQIRSGGSPVTHIGTFAELGLGVVDNNGNGAR 283
QY 61 VQRVGAPAPASLGISTGDIYTAVDGAPINSATRAMDALNGHHPGDIYSTWQTKSGGGR 120
Db 284 VQRVGAPAPASLGISTGDIYTAVDGAPINSATRAMDALNGHHPGDIYSTWQTKSGGGR 343
QY 121 TGNVTLAEGRPA 132
Db 344 TGNVTLAEGRPA 355
RESULT 2
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999
C:Accession: S47170
R: Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
Submitted to the EMBL Data Library, June 1993
A: Description: Isolation and characterisation of a 34Kda protein of Mycobacterium par
A: Reference number: S47170
A: Accession: S47170

A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-361 <Cam>
A:Cross-references: EMBL:Z23092; NID:g505550; PIDD:CAA80638.1; PID:g505551
C:Superfamily: protease hnoB

| | | | | |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match | 70.3%; | Score 474.5; | DB 2, | Length 361; |
| Best Local Similarity | 69.7%; | Pred. No. 3.9e-32; | | |
| Matches 92; | Conservative 17; | Mismatches 22; | Indels 1; | Gaps 1 |

| | | | |
|----|-----|---|-----|
| QY | 1 | TAASNPDSGGGGGFAIPIDQAAIIGOLISGGSGSTVNHGPAFAGLGVNDNGGAR | 60 |
| | | : : : : : : : : : : | |
| Db | 231 | TAATSTSYMS-GGCGFAPIDGRAMAANQISGSGSNVNHGPTAFGLGVTDNNGGAR | 289 |
| QY | 61 | VORVVGSPAPASLSTGSDVTVAVDGAPINETAAMALNGHHNGDVISTWQKSSGSTR | 120 |
| | | : : : : : : : : : | |
| Db | 290 | VQRVNVNTPPAAAGAGAPBDVITGVDPINAKISMTSEVLPRHHNGCDIIVNHRVLDGGR | 349 |
| QY | 121 | TGNVTLLAGCPRA | 132 |
| | | : | |
| Db | 350 | TANITLLAGCPRA | 361 |

```

RESULT      3
A:87242
probable secreted serine proteinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87242
R:COLE, S.T.; Elgimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hsiao, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, C.;
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Ruter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Scott,
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: A87242
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1:354 <STO>
A:Cross-references: GB:AL450380; NID:g13093863; PIDN:CAC32191.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML2659

```

| | | | | |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match | 68.1%; | Score 459.5; | DB 2; | Length 354; |
| Best Local Similarity | 66.7%; | Pred. No. 6.7e-31; | | |
| Matches 88; | Conservative 17; | Mismatches 26; | Indels 1; | Gaps 1 |

[illegible]

RESULT 4
C70821
probable serine proteinase Rv0983 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70821
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: G10921

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-464 <COL>
A:Cross-references: GB:AL021999; GB:AL123456; NTD:33261538; PIDN:CAA17582.1; PID:el25
A:Experimental source: strain H37Rv
A:Genetics: RV0983

| | | | | |
|-----------------------|--------|--------------------|-------|----------------------------------|
| Query March | 29.6%; | Score 199.5; | DB 2; | Length 464; |
| Best Local Similarity | 38.9%; | Pred. No. 3.2e-09; | | |
| Matches | 51; | Conservative | 21; | Mismatches 50; Indels 9; Gaps 3; |

| | | | | | |
|----|-----|---|---|--------|-----|
| QY | 1 | TAASDNFQLSQG | -GGGFAIPFGQAMALAGQIRSGGSPVTHIGPFAELGLGV | -DNGCN | 57 |
| | | : : | : | : : | |
| Db | 337 | TLGADSDAAGSGSIGLFAIPVDQAKRIADLELISTGKA | -----SHASLGQVITNDKOTL | | 390 |
| QY | 58 | GARYQRVYSGAPASLIGISTGDYITAVDCAPINSTATMAADLNHNGHDYISTVNTQKSG | | | 117 |
| | | : : | : : | : | |
| Db | 391 | CAKVEVAVAGGAANAGVPRGVVTKVDDRPINSADALVAAVRSKAGCATVALLTFDDPSG | | | 450 |
| QY | 118 | GIRTGNVTILAE | | | 128 |
| | | : : | | | |
| Db | 451 | GSRTVQVTLTK | | | 461 |

RESULT 5
H86930
probable secreted serine proteinase [imported] - *Mycobacterium leprae*
C:Species: *Mycobacterium leprae*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H86930
R:Cole, S.T.; Eigimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holton
eam, M.A.; Rutherford, K.M.
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H86930
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-382 <STO>
A:Cross-references: GB:AL450380; NID:g13092536; PIDN:CAC29684.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML0176
C:Superfamily: *Escherichia coli* trypsin-like proteinase

| | | | | |
|--------------------------|-------|--------------------|-----------|-------------|
| Query Match | 27.9% | Score 188.5; | DB 2; | Length 382; |
| Best Local Similarity | 39.8% | Pred. No. 2.1e-08; | | |
| Matches 51; Conservative | 19; | Mismatches 51; | Indels 7; | Gaps 3; |

[illegible]

RESULT 6
T45448
probable serine proteinase (EC 3.4.21. -) MLCB373.28 [similarity] - Mycobacterium leprae

Db 302 GAADAGLRGVDYLVKIGDTITTTSLSEALASMRPGDKVTYV-TRGKREHTAEVTLG 360
 QY 128 E 128
 Db 361 E 361

RESULT 10

H86891

exported serine proteinase (EC 3.4.21.-) [imported] - Lactococcus lactis subsp. lactis
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C:Accession: H86891

R:Polotkin, A.; Winkler, P.; Mager, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrlich
 Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
 A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: H86891

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-408 <STO>

A:Cross-references: GB:AE005176; PID:g12725194; PIDN:AAK06234.1; GSPDB:GN00146

A:Experimental source: strain 114403

C:Genetics:

A:Gene: htra

C:Keywords: hydrolase; serine proteinase

Query Match 19.8%; Score 133.5; DB 2; Length 408;
 Best Local Similarity 31.1%; Pred. No. 0.00083;

Matches 41; Conservative 20; Mismatches 48; Indels 23; Gaps 5;

QY 13 GGGFAIPGQAMAIAGQIRSGGSPVTHIGPFLGLGVDD-----NNGN----- 57
 Db 270 GGGFAIPGQAMAIAGQIRSGGSPVTHIGPFLGLGVDD-----NNGN----- 57

QY 58 -GARVORVGSAPASLGLSTGDTTAVDGPINSATAMADALNGHHHPEDVISVTWQTKS 116
 Db 324 GGVVSVSVSGLEPAASAGLAKADVDITKVDVTSSTDLQSLALSHINDIVKVTYY-RD 382

QY 117 GGTGTGNTVLAE 128
 Db 383 GKSNTADVTKSK 394

RESULT 11

E75357

probable periplasmic serine proteinase Do - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: E75357

R:White, O.; Eelsen, J.A.; Heideberg, J.F.; Hickey, E.R.; Peterson, J.D.; Dodson, R.J.;
 S.; Shults, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896

A:Accession: E75357

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <WMT>

A:Cross-references: GB:AE002017; GB:AE000513; NID:g6459527; PIDN:AAF11312.1; PID:g645953
 A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1756

A:Map position: 1

C:Superfamily: proteinase hhob

Query Match 19.4%; Score 131; DB 2; Length 441;
 Best Local Similarity 29.0%; Pred. No. 0.0015;
 Matches 47; Conservative 24; Mismatches 43; Indels 48; Gaps 6;

QY 6 NFQISGGQ-----GFAIPGQAMAIAGQIRSGGSPVTHIGPFLGLGVDDN---- 55
 Db 277 NQIILTGAGGAGAGGAFIPINTVKRLPOLQAGKG-----GVSPSPISGVFSDLSL 330

QY 56 -----GNGARVORVGSAPASLGLSTGDTTAVDGPINSATAMADALNGHHHPEDVISVTWQTKS 126
 Db 331 PPQOLKAAGLPSSGALLOKVPSPSPAAAGLRGNNKSLSPASQGSISTDGLITAV 330

QY 85 DGAPINSATAMADALNGHHHPEDVISVTWQTKSGGTFTGNTL 126
 Db 391 MGPLEDAGSLQFAVLATGEGQPLRTVR-RGKTRREVETL 431

RESULT 12

A82581

periplasmic proteinase XF2241 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A82581

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A82581

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-514 <STIM>

A:Cross-references: GB:AE004037; GB:AE003849; NID:g9107394; PIDN:AAF5040.1; GSPDB:GN
 A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Accencio, M.; Alvarenga, R.
 Bioness, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
 as-Neto, E.; Docena, C.; El-Doroty, H.; Facinelli, A.P.; Ferreira, A.D.S.
 submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr
 J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, F.R.; da Silva Jr., W.A.; da Silv
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2241

C:Superfamily: Helicobacter serine proteinase

Query Match 19.0%; Score 128.5; DB 2; Length 514;
 Best Local Similarity 38.7%; Pred. No. 0.0028;
 Matches 43; Conservative 11; Mismatches 46; Indels 11; Gaps 4;

QY 4 SDNEQLSQG--GGFAIPGQAMAIAGQIRSGG-----GSPVTHIGPFLGLGVDDN 54
 Db 262 SQPSASGAGYMSIFAIPINIAAEQIRKTKGVORSMLGVETIGPDLAKAGGLGPS 321

QY 55 NGNGARVORVGSAPASLGLSTGDTTAVDGPINSATAMADALNGHHHPG 105
 Db 322 --KCALVNNIPHPSPAKAGIEVDVIRSVNGKVISSFDPLPLIGMPPG 370

RESULT 13

F83550

serine proteinase MucD precursor PA0766 [imported] - Pseudomonas aeruginosa (strain P
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83550

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
 ; Lbray, S.; Olson, M.V.
 Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337
 A:Accession: F8350
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-474 <STO>
 A:Cross-references: GB:AE004511; GB:AE004091; NID:g9946646; PIDN:AAG04155.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: mucD; PA0766

Query Match 18.1%; Score 122.5; DB 2; Length 474;
 Best Local Similarity 30.4%; Pred. No. 0.008;
 Matches 42; Conservative 22; Mismatches 51; Indels 23; Gaps 5;
 QY 4 SDNFOLSG--GQFAIPDIGAMAIAGQIRSGGSPVYHIGPT--AF-----LIGVYDNNNGARVQR 63
 DB 233 SGIFFRSGGFMSFAPIDVAVLNVADLKKAG--KVSRG--WLGVIQEVNKKDLAES 286
 QY 56 -----GNCARVQRYVGSAPASLIGSTGDIYTAVDGAPINSATAMADALNGHHPGDVTSV 110
 DB 287 FGIDKPSGALVAQLVEBGPAPKAGLQVGDVLTSLNGSINSADLPHLVGMKFGDKINL 346
 QY 111 ----TWQTKSGGTRTGNV 124
 DB 347 DVIRNGQRKSLSMVAGSL 364

RESULT 14 735866
 probable integral membrane protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence.revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: J35866
 R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
 submitted to the EMBL Data Library, April 1999
 A:Reference number: Z21591
 A:Accession: J35866
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-203 <SAU>
 A:Cross-references: EMBL:AL049727; PIDN:CABA1567.1; GSPDB:GN00070; SCQDB:SC9B1.21
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCQDB:SC9B1.21

Query Match 17.7%; Score 119.5; DB 2; Length 203;
 Best Local Similarity 32.1%; Pred. No. 0.0057;
 Matches 36; Conservative 15; Mismatches 56; Indels 5; Gaps 1;
 QY 12 GGCGFAIPDIGAMAIAGQIRSGGSPVYHIGPT-----AFGLGVNDNNGARVQRVYG 66
 DB 80 GGAGEGATGTATGTGCAHPHRSAGRSFSPAPAGATLGEAVDDEKPGARVGVHV 139
 QY 67 SAPASLIGSTGDIYTAVDGAPINSATAMADALNGHHPGDVTSVWQTKSGG 118
 DB 140 PGPGYAAGLVAGVDLAVGTTRVDSATDLAAHVARAGPGKEVKLTIVRHRSRG 191

RESULT 15 F87590
 serine proteinase HtrA [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence.revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: F87590
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolod
 n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87590
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-530 <STO>
 A:Cross-references: GB:AE005673; NID:g13424352; PIDN:AAK24722.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC2758
 C:Superfamily: Helicobacter serine proteinase

Query Match 17.4%; Score 117.5; DB 2; Length 530;
 Best Local Similarity 31.6%; Pred. No. 0.023;
 Matches 42; Conservative 18; Mismatches 50; Indels 23; Gaps 5;
 QY 13 GQFAIPDIGAMAIAGQIRSGGSPVYHIGPT--AF-----LIGVYDNNNGARVQR 63
 DB 283 GIGFALPAEYVAGVAKQLVNGVVRGYIGVSIMARNAEMAEALGMSDV---KGLAVAS 338
 QY 64 VGSAPASLIGSTGDIYTAVDGAPINSATAMADALNGHHPGDVTSV-----WQT 114
 DB 339 VVPGGPAKAKGLPDDILVAVNGVKISDSSELTREVSKARPGETIKVSIIRDGKPRIVDV 398
 QY 115 KSGGTRTGNVTLA 127
 DB 399 KS-GTPRESSLA 410

Search completed: August 6, 2002, 08:16:53
 Job time: 253 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:15:10 ; Search time 10.04 Seconds
(without alignments)
509.062 Million cell updates/sec

Title: US-09-684-215a-4

Perfect score: 675
Sequence: 1 TAASDNFOLSGGCGFAIP1.....QTKSGGTGTGNVTLAEGPPA 132

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|---------------------|
| 1 | 133.5 | 19.8 | 408 | HTRA_LACLA | Q91A06 Lactococcus |
| 2 | 125 | 18.5 | 413 | HTRA_LACHE | Q9Z4H7 Lactobacilli |
| 3 | 118 | 17.5 | 504 | DEGP_RHIME | Q52894 Rhizobium m |
| 4 | 113 | 16.7 | 513 | DEGP_BRUAB | Q44597 Brucella ab |
| 5 | 106.5 | 15.8 | 355 | DEGS_ECOLI | P31137 Escherichia |
| 6 | 106 | 15.7 | 455 | DEGO_ECOLI | P39089 Escherichia |
| 7 | 105 | 15.6 | 475 | DEGP_SALTY | P26982 Salmonella |
| 8 | 102.5 | 15.2 | 474 | DEGP_ECOLI | P09376 Escherichia |
| 9 | 101.5 | 15.0 | 503 | DEGP_BARE | P54925 Bartonella |
| 10 | 97 | 14.4 | 497 | DEGP_CHLTR | P18584 Chlamydia t |
| 11 | 94.5 | 14.0 | 497 | DEGP_CHLNU | Q9P197 Chlamydia m |
| 12 | 94.5 | 14.0 | 957 | Y278_MYCTU | P56877 Mycobacteri |
| 13 | 92.5 | 13.7 | 466 | HTON_HAEIN | P45179 Haemophilus |
| 14 | 92 | 13.6 | 448 | DEGP_ARATH | Q91U10 Arabidopsis |
| 15 | 92 | 13.6 | 478 | DEGP_BUCAI | P57332 Buchnera ap |
| 16 | 89.5 | 13.3 | 905 | Z03_MOUSE | Q9QXY1 Mus musculu |
| 17 | 87 | 12.9 | 864 | ELIS_RAT | Q99372 Rattus norv |
| 18 | 86 | 12.7 | 762 | P15_CHICK | Q98917 Gallus gall |
| 19 | 86 | 12.7 | 2249 | OMPA_RICRI | P15931 Rickettsia |
| 20 | 85 | 12.6 | 515 | Y140_MYCTU | Q50534 Mycobacteri |
| 21 | 85 | 12.6 | 2021 | OMPA_RICCN | Q52657 Rickettsia |
| 22 | 83 | 12.3 | 488 | DEGP_CHLNU | Q92650 Chlamydia p |
| 23 | 83 | 12.3 | 510 | MURF_MYCTU | Q06220 Mycobacteri |
| 24 | 82 | 12.1 | 389 | HMXU_DROME | P02834 Drosophila |
| 25 | 81.5 | 12.1 | 219 | VG36_BPT2 | P07066 Bacterioph |
| 26 | 81.5 | 12.1 | 619 | ELF1_HUMAN | P22519 Homo sapien |
| 27 | 81.5 | 12.1 | 714 | CDG1_PALMA | P04830 Penicillium |
| 28 | 81 | 12.0 | 354 | ODPA_ZYMO | O66112 Zymomonas m |
| 29 | 81 | 12.0 | 430 | TPSN_CHICK | O73895 Gallus gall |
| 30 | 80 | 11.9 | 382 | YN28_MYCTU | P17184 Mycobacteri |
| 31 | 79.5 | 11.8 | 292 | PPNK_ECOLI | P58057 Escherichia |
| 32 | 79.5 | 11.8 | 292 | PPNK_ECOLI | P37766 Escherichia |
| 33 | 79.5 | 11.8 | 1286 | AIDA_ECOLI | O03155 Escherichia |

| | | | | | | |
|----|------|------|-----|---|------------|--------------------|
| 34 | 79 | 11.7 | 253 | 1 | HMXU_DROME | P05048 drosophila |
| 35 | 79 | 11.7 | 257 | 1 | ECHE_MYCLE | O07137 mycobacteri |
| 36 | 78.5 | 11.6 | 351 | 1 | HYPE_ALCEU | P31905 atc11genes |
| 37 | 78 | 11.6 | 159 | 1 | 19KD_MYCTU | P11572 mycobacteri |
| 38 | 78 | 11.6 | 801 | 1 | Y747_MYCTU | O53810 mycobacteri |
| 39 | 77.5 | 11.5 | 478 | 1 | DEGP_BUCAI | O85291 buchnera ap |
| 40 | 77.5 | 11.5 | 479 | 1 | XVYP_LACPE | P96792 lactobacill |
| 41 | 77.5 | 11.5 | 933 | 1 | Z03_HUMAN | O95049 homo sapien |
| 42 | 77 | 11.4 | 396 | 1 | YD61_MYCTU | O11031 mycobacteri |
| 43 | 77 | 11.4 | 543 | 1 | YUP1_DROME | P16375 drosophila |
| 44 | 77 | 11.4 | 746 | 1 | YUP2_DROME | P16376 drosophila |
| 45 | 77 | 11.4 | 778 | 1 | YQ34_MYCTU | P71933 mycobacteri |

ALIGNMENTS

RESULT 1
HTRA_LACLA STANDARD; PRT; 408 AA.
ID HTRA_LACLA
AC Q9LA06;
DT 16-OCT-2001 (Rel. 40, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DR 01-MAR-2002 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-) (HtraLl).
GN HTRA OR IL2136
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
ON NCBI_TaxID=1360;
RX [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=IL1403;
RX MEDLINE=20177820; PubMed=10712686;
RA Poquet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;
RT "Htra is the unique surface housekeeping protease in Lactococcus
RT lactis and is required for natural protein processing.";
RL Mol. Microbiol. 35:1042-1051(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Mauger S., Jallion O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: Degrades abnormal exported proteins. Needed for the pro-
CC peptide processing of a natural pro-protein and for maturation of
CC A native protein. Responsible for the housekeeping of exported
CC proteins.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF155705; AAF61294.1; -
CC EMBL: AE006442; AAK06234.1; -
CC InterPro: IPR001478; PDZ
CC InterPro: IPR001940; ProteaseZC.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00595; PDZ; 1.
CC Pfam: PF00089; Trypsin; 1.
CC PRINTS: PR00834; PROTEASES2C.
CC SMART: SMO0228; PDZ; 1.

DR PROSITE: PS50106; PDZ; 1.
 KM Hydrolase; Serine protease; Transmembrane; Complete proteome.
 FT TRANSMEM 26
 FT DOMAIN 8 284 POTENTIAL.
 FT ACT_SITE 302 383 CATALYTIC.
 FT ACT_SITE 127 127 PDZ.
 FT ACT_SITE 157 157 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 239 239 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 239 239 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 408 AA; 41648 MW; 581B90B5A7DF851 CRC64;

Query Match
 Best Local Similarity 31.1%; Score 133.5; DB 1; Length 408;
 Matches 41; Conservative 20; Mismatches 48; Indels 23; Gaps 5;

OY 13 GCGFAIPGQAMAIAGQIRSGGSPVTHIGPAFLGLGYVD-----NNGN----- 57
 DB 270 GIGFAIPSDVNIINKLADG----KISRPA-IGIRVVDLSQLSTNDSSQIKLPSSVT 323
 OY 58 -GARVQRVVGSAPAAASLIGSTGDTAVDAPINSATAMADALNGHHPGDIYSVTQTKS 116
 DB 324 GGVVYVSQGLPAAASLIGADYITKGDPTAVTSTDLQSLYSHINDYVAVTY-RD 382

OY 117 GGTGTGNTVLAEL 128
 DB 383 GKSNFADVKLSK 394

RESULT 2

HTRA_LACHE STANDARD; PRT; 413 AA.
 ID HTRA_LACHE
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine protease do-like htra (PC 3.4.21.-).
 GN HTRA.

OS Lactobacillus helveticus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1587;
 RN (1)
 RP SEQUENCE FROM N.A.

RC STRAIN=53/7;
 RX MEDLINE=99047559; PubMed=9829922;
 RA Smets A., Varmann P.K., Palva A.M.;

RT "Molecular characterization of a stress-inducible gene from

RT Lactobacillus helveticus";

RL J. Bacteriol. 180:6143-6153(1998).

CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE

CC DEGP/DEGC/DEGS FAMILY.

CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AJ005672; CAA06668.1; -

DR MEROPS: S01.273; -

DR InterPro: IPR001478; PDZ.

DR Pfam: PF00089; trypsin; 1.

DR PRINTS: PR00834; PROTEASES2C.

DR SMART: SM00228; PDZ; 1.

DR PROSITE: PS50106; PDZ; 1.

KW Hydrolase; Serine protease; Transmembrane.
 FT TRANSMEM 20 40 POTENTIAL.

FT DOMAIN 141 302 CATALYTIC.
 FT DOMAIN 305 401 PDZ.
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 413 AA; 42647 MW; B16B677991C88707 CRC64;

Query Match
 Best Local Similarity 30.5%; Score 125; DB 1; Length 413;
 Matches 36; Conservative 16; Mismatches 52; Indels 14; Gaps 3;

OY 8 QLSQSG---GCGFAIPGQAMAIAGQIRSGG--GSPVTHIGPAFLG-----LGVD 53
 DB 279 QSSDGTSGVSEMAFAIPSENVVTVNELVKKITTPOGLGVAVIALGQIPCYRSRLTKS 338
 OY 54 NNGGARVQRVVGSAPAAASLIGSTGDTAVDAPINSATAMADALNGHHPGDIYSVT 111
 DB 339 NLKNGYIAFVSRRNGSAANAGIKSGDYITKGDVKKVEDVASLSHLYSHKVGDTVNT 396

RESULT 3

DEGP_RHIME STANDARD; PRT; 504 AA.
 ID DEGP_RHIME
 AC 052894;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (PC 3.4.21.-).
 GN DEGP1 OR DEGP OR R01021 OR SMC02365.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.

RC STRAIN=1021;
 RX MEDLINE=96146524; PubMed=8550509;
 RA Glazebrook J., Ichige A., Walker G.C.;

RT "Genetic analysis of Rhizobium meliloti bacA-phoA fusion results in

RT identification of degP: two loci required for symbiosis are closely

RT linked to degP";

RL J. Bacteriol. 178:745-752(1996).

CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: U31512; AAC43669.1; ALT INIT.

DR EMBL: AL591785; CAC45593.1; -

DR MEROPS: S01.273; -

DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.


```

RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamodis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kudara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=88105815; PubMed=332223;
RA Vogel R.F., Entlian K.-D., Mecke D.;
RT "Cloning and sequence of the mdh structural gene of Escherichia coli
RT coding for malate dehydrogenase.";
RL Arch. Microbiol. 149:36-42(1987).
RN [7]
RP IDENTIFICATION.
RA Bazan J.F., Fletterick R.J.;
RT "Structural and catalytic models of trypsin-like viral proteases.";
RL Semin. Virol. 1:311-322(1990).
CC -1- SUBCELLULAR LOCATION: Periplasmic (potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGP/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43993.1; -
DR EMBL; U32495; AAC44006.1; -
DR EMBL; U18997; AAA58037.1; -
DR EMBL; AE000402; AAC76267.1; -
DR EMBL; AE005551; AAG58363.1; -
DR EMBL; AP002564; BAB37531.1; -
DR EMBL; M24777; -; NOT_ANNOTATED_CDS.
DR MEROPS; S01.275; -
DR EcoGene; EG11652; hohB.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF000595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.

```

```

DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Serine protease; periplasmic; signal; complete proteome.
FT SIGNAL 1
FT CHAIN 29 355
FT DOMAIN 281 326
FT ACT_SITE 96 96
FT ACT_SITE 126 126
FT ACT_SITE 201 201
FT CONFLICT 253 253
FT CONFLICT 307 307
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;

Query Match 15.8%; Score 106.5; DB 1; Length 355;
Best Local Similarity 31.2%; Pred. No. 0.092;
Matches 44; Conservative 10; Mismatches 64; Indels 23; Gaps 5;

OY 6 NFOLSGGQ-----GFAIP-----IGGAMLAGIRSGGSPYHIGPTFLGL 49
DB 219 SFKMSDGEFPEIGTFAIFPOLATKIMDKLIRGRVIRGTYIGGR---ELAPLHAQGG 274
OY 50 GVVNDNGNGARQVYVGSAPASLIGSTGDTITAVDGPINSATAMADALNGHPDVIS 109
DB 275 GI--DQLGIVNEVSPDGPAAAGIQVNDLISVKNKPAISALETMDQVAELRPSVIP 332
OY 110 VTMQTSGGTGRGNTVLAQSP 130
DB 333 VV-VMRDKQLTLQVITQIEP 352

RESULT 6
DECO_ECOLI
ID DECO_ECOLI STANDARD; PRT; 455 AA.
AC P39099;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protease degp precursor (EC 3.4.21.-).
GN DEGP OR HHOA OR B3234.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA Bass S., Gu O., Goddard A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 28-32.
RC STRAIN-K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RT Weller P.R., Sauer R.T.;
RT "Characterization of degp and degs, Escherichia coli genes encoding
RT homologs of the Degp protease.";
RL J. Bacteriol. 178:1146-1153(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGP/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U15661; AAC3992.1; -
 DR EMBL: U32495; AAC4005.1; -
 DR EMBL: U18997; AAA58036.1; -
 DR EMBL: AE000402; AAC76266.1; -
 DR MEROPS: S01.274; -
 DR SWISS-2DPAGE: P39099; COLI.
 DR Ecocore: EG12612; dego.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 DR HydroLase: Serine protease; periplasmic; signal; complete proteome.
 KW SIGNAL
 FT CHAIN 1 27
 FT DOMAIN 28 435
 FT DOMAIN 258 349
 FT ACT_SITE 355 447 PDZ 1.
 FT ACT_SITE 109 109 PDZ 2.
 FT ACT_SITE 139 139 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;

Query Match 15.7%; Score 106; DB 1; Length 455;
 Best Local Similarity 26.5%; Pred. No. 0.13;
 Matches 36; Conservative 19; Mismatches 41; Indels 40; Gaps 5;

QY 13 GGGFAIPGQAMIA-----GGIRSGGSPVNIIGPTAFGLGVNDNGN----- 57
 DB 241 GIGFAIPSNMAYFLAOQLIDFGEIKRG-----DLGIKGTMSADIAKAFMLDV 288
 QY 58 --GARQRYVGSAPASLSIGSTDVITAVDGAIPNSATAMADLNGHHPGDVISTVWTK 115
 DB 289 QKRAFYSEVLPGGSGAKAGKADITSLNGKPLNSFALRLSRITATPEPTG-----KVK 342
 QY 116 SGGTRTG-----NVTL 126
 DB 343 LGLLRNCKPLEVEYTL 358

RESULT 7
 DEGP_SALTY STANDARD; PRT; 475 AA.
 ID DEGP_SALTY
 AC P26982;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Protease do precursor (EC 3.4.21.-)
 GN DEGP OR HTA OR PTD OR STM0209.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C5;
 RX MEDLINE=91251770; PubMed=1645840;
 RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
 RA Ali T., Miller I., Hormaeche C.;
 RT "The role of a stress-response protein in Salmonella typhimurium
 RT virulence."
 RL Mol. Microbiol. 5:401-407(1991).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvey E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2."
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
 CC SPECIFICITY WITH HROA/DEGP.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X54548; CAA38420.1; -
 DR EMBL: AE008704; AAL19173.1; -
 DR PIR: S15337; S15337.
 DR PIR: S21327; S21327.
 DR MEROPS: S01.273; -
 DR StyGene: SG10173; degp.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 DR HydroLase: Serine protease; Heat shock; periplasmic; signal;
 KW Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 475
 FT DOMAIN 281 372
 FT DOMAIN 378 467 PDZ 1.
 FT ACT_SITE 132 132 PDZ 2.
 FT ACT_SITE 162 162 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 475 AA; 49315 MW; 86E65BF3C1A289F CRC64;

Query Match 15.6%; Score 105; DB 1; Length 475;
 Best Local Similarity 28.1%; Pred. No. 0.16;
 Matches 38; Conservative 19; Mismatches 44; Indels 34; Gaps 4;

QY 13 GGGFAIP-----IGQAMINQIRSGGSPVNIIGPTAFGLGVNDNGN----- 57
 DB 264 GIGFAIPSNMAYKNTLSQWVEYGOVVRG-----ELGIMGTLSLSELAANKV 309
 QY 58 ----GARQRYVGSAPASLSIGSTDVITAVDGAIPNSATAMADLNGHHPGDVISTVWTK 113
 DB 310 DAQRGAFVSQVMPNSAKAGIKAGDVITSLNGKPISSFALRAQVGMIPGSKISTL-GL 368
 QY 114 TKSGGTRTGNVTLAE 128
 DB 369 LREGKAITVNIETLQ 383

RESULT 8
 DEGP_ECOLI

ID DEGP_ECOLI STANDARD: PRT; 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htra gene of Escherichia coli: a sigma 32-independent mechanism of heat-inducible transcription.";
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E., Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D., Davis R.W.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL [5]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grothbeck E.U., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsuno E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90323597; PubMed=2165018;

RA Quirk S., Bhattachar S.K., Bessman M.J.;
RT "Primary structure of the deoxyguanosine triphosphate
RL triphosphohydrolase-encoding gene (dgt) of Escherichia coli.";
RN Gene 89:13-18(1990).
RN [8]
RP SEQUENCE OF 1-16 FROM N.A.
RX MEDLINE=90207273; PubMed=2157212;
RA Wurgler S.M., Richardson C.C.;
RT "Structure and regulation of the gene for dGTP triphosphohydrolase from Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
RN [9]
RP CHARACTERIZATION AND SEQUENCE OF 27-39.
RX MEDLINE=90202693; PubMed=2180903;
RA Lipinska B., Zyllicz M., Georgopoulos C.;
RT "The htra (degp) protein, essential for Escherichia coli survival at high temperatures, is an endopeptidase.";
RL J. Bacteriol. 172:1791-1797(1990).
RN [10]
RP IDENTITY OF HTRA AND PROTEASE DO.
RX MEDLINE=91222240; PubMed=2025286;
RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J., Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
RT "Protease Do is essential for survival of Escherichia coli at high temperatures: its identity with the htra gene product.";
RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE. INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGP.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT TEMPERATURES ABOVE 42 DEGREES CELSIUS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE DEGP/DEGO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36536; AAA23994.1; -;
DR EMBL: X12457; CAA30997.1; -;
DR EMBL: D26562; CAB20280.1; -;
DR EMBL: AE000125; AAC73272.1; -;
DR EMBL: U70214; AAB08591.1; -;
DR EMBL: AE005192; AAG54465.1; -;
DR EMBL: AF002550; BAB33588.1; -;
DR EMBL: M29955; AAA23717.1; -;
DR EMBL: M31772; AAA23680.1; -;
DR PIR: S01899; S01899.
DR PIR: B35993; B35993.
DR MEROPS: S01.273; -;
DR SWISS-2DPAGE: P09376; COLI.
DR EcoGene: EG10463; degp.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR Pfam: PF00595; PDZ_2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ_2.
DR PROSITE: PS50106; PDZ_2.
KW Hydrolyase; Serine protease; Heat shock; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 474
FT DOMAIN 280 371
FT PROTEASE DO.
FT PDZ 1.

```

FT DOMAIN 377 466 PDZ 2.
FT ACT_SITE 131 131 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 161 161 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 236 236 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 10 10 A -> R (IN REF. 1, 7 AND 8).
FT CONFLICT 46 46 E -> Q (IN REF. 7).
FT CONFLICT 192 192 A -> G (IN REF. 1).
FT CONFLICT 467 474 STYLYMO -> RHLFVNVAIVSLNPLKTKGRGSPFYL (IN
FT SEQUENCE 474 AA; 49354 MW; 5482E59674B6D5F CRC64;
SQ QUERY MATCH 15.2%; Score 102.5; DB 1; Length 474;
Best Local Similarity 30.4%; Pred. No. 0.25; Mismatches 33; Gaps 3;
Matches 31; Conservative 14; Indels 24;
QY 13 GQGFAIP-----TGOAMAIAGQIRSGGSPYVHIGTFAELGICVDNNGN----- 57
DQ 263 GIGFAFVSQVLPNNSAKAGIKAGDVITSLNGKPISSFAL 308
OY 58 -----GARQQRVYGSAPASLGIISTGVDITANDGAPINSATAM 95
DB 309 DAORGAFVSVQVLPNNSAKAGIKAGDVITSLNGKPISSFAL 350
RESULT 9
DEGP_BARNE STANDARD; PRT; 503 AA.
AC P54925;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
DE (Antigen htra).
DE DEGP OR HTRA.
GN Bartonella henselae (Kochallimaea henselae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=38323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOUSTON-1.
RX MEDLINE=94239828; PubMed=8027347;
RA Anderson B., Sims K., Regniery R., Robinson L., Schmidt M.J.,
RA Goral S., Hager C., Edwards K.;
RT "Detection of Kochallimaea henselae DNA in specimens from cat scratch
RT disease patients by PCR.";
RU J. Clin. Microbiol. 32:942-948(1994).
CC -I- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -I- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L20127; AAA97430.1; -.
DR MEROPS: S01.273; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001544; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROPEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR Hydrolase; Serine protease; Periplasmic; Signal.
KW SIGNAL 1 18
FT POTENTIAL.

```

```

FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-
FT DOMAIN 286 357 LIKE.
FT ACT_SITE 419 466 PDZ 1.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SO SEQUENCE 503 AA; 54114 MW; 6CD9F4743282AF9E CRC64;

Query Match 15.08; Score 101.5; DB 1; Length 503;
Best Local Similarity 26.5%; Pred. No. 0.32; Indels 41; Gaps 6;
Matches 39; Conservative 15; Mismatches 52;

OY 7 FOLSGGGG---FAIPDIGAMAIAGQIRSGGSPVTHIGTPAIFGLGYVDNNGNGARQVY 64
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 266 FSPSGGVNGVIAFAIPATANEVQOL-----TEKGLVQGMGVQIDQV 309
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 65 -----VG-----SAPAASLGISTGDTVAVDGAFINSATAMADALNGHP 104
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 310 TKEISDSTIGLKEAGKALITDPLKGPAAKAGIKADVIISVGEKINDVRLAKRIANMSP 369
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 105 GDVIVS-ITWQTRSGGTRTGNTLAEGP 130
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 370 GETVTLGVW--KSGKEENIKVKLDSP 394
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
DEGP_CHLTR STANDARD: PRT; 497 AA.
ID DEGP_CHLTR
AC P18584; 084830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
DE immunogenic protein) (SK59).
DE DEGP OR HTRA OR CT823.
OS Chlamydia trachomatis.
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=SEROVAR I2;
RC MEDLINE=90337348; PubMed-2379836;
RA Kane S., Weinstein Y., Sarov I.,
RT "Cloning, characterization and sequence of a novel 59-kDa protein of
RL Chlamydia trachomatis.";
RN Gene 90:61-67(1990).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=D/TW-3/CX;
RC MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGP/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -1- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL: AE001355; AAC68420.1; -
DR FMBL: M31119; AAA23116.1; -
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR00126; Ser_proteas_V8.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS0106; PDZ; 2.
KW Hydrolyase; Serine protease; Signal; Antigen; Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 497
FT DOMAIN 128 289
FT DOMAIN 290 381
FT DOMAIN 394 485
FT ACT_SITE 143 143
FT ACT_SITE 173 173
FT ACT_SITE 247 247
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31BB84A38BA CRC64;

```

```

Query Match 14.4%; Score 97; DB 1; Length 497;
Best Local Similarity 29.9%; Pred. No. 0.73;
Matches 38; Conservative 18; Mismatches 57; Indels 14; Gaps 3;

```

```

QY 13 GGGFAIPITGMAAIAAGQIRSGGSPYVHIGPTAFGLGVND-----NGNGARVOR 63
DB 274 GIGFAIPISLMKRVYDQLSDGVTRGFLGVT----LQPIDSELAACYKLEKYGALVTD 329
QY 64 VGSAPASLSIGTSDVITVAVDGAIPNSATAMADALNGHPGVISVTWQTKSGGRTGN 123
DB 330 VVGSPAEKAGLRQEDVIVAYNGKEVESLSALRNATISLMMPGRV-VLKIRGKRTIEP 388
QY 124 VTLAEGP 130
DB 389 VVTVOIP 395

```

```

RESULT 11
DEGP_CHLMT STANDARD; PRT; 497 AA.
AC G9P197;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR TC0210.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaeae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uetereck T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEQ/DEGS FAMILY.
CC - SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: AE002288; AAF39082.1; -
DR TIGR: TC0210; -
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR00126; Ser_proteas_V8.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: SM00228; PDZ; 2.
DR PROSITE: PS0106; PDZ; 2.
KW Hydrolyase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 497
FT DOMAIN 128 289
FT DOMAIN 290 381
FT DOMAIN 394 485
FT ACT_SITE 143 143
FT ACT_SITE 173 173
FT ACT_SITE 247 247
SQ SEQUENCE 497 AA; 53294 MW; B765F350AC66BBF CRC64;

```

```

Query Match 14.0%; Score 94.5; DB 1; Length 497;
Best Local Similarity 30.4%; Pred. No. 1.2;
Matches 31; Conservative 15; Mismatches 43; Indels 13; Gaps 2;

```

```

QY 13 GGGFAIPITGMAAIAAGQIRSGGSPYVHIGPTAFGLGVND-----NGNGARVOR 63
DB 274 GIGFAIPISLMKRVYDQLSDGVTRGFLGVT----LQPIDSELAACYKLEKYGALVTD 329
QY 64 VGSAPASLSIGTSDVITVAVDGAIPNSATAMADALNGHPG 105
DB 330 VVGSPAEKAGLRQEDVIVAYNGKEVESLSALRNATISLMMPG 371

```

```

RESULT 12
Y278_MYCTU STANDARD; PRT; 957 AA.
ID Y278_MYCTU
AC P56877;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Hypothetical PE-PGRS family protein RV0278c precursor.
GN RV0278c OR MT0291 OR MT035.06c.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsley T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.,
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AL021930; CAA17353.1; -
 DR EMBL: AE006936; AAK44511.1; ALT_INIT.
 DR TIGR: MT0291; -
 DR TubercuList: Rv0278c; -
 DR InterPro: IPR000084; PE.
 DR Pfam: PF00934; PE; 1.
 DR Hypothetical protein; Repeat; Signal; Complete proteome.
 KM SIGNAL 1 30 POTENTIAL.
 FT CHAIN 1 957 HYPOTHEICAL PE-PGRS FAMILY PROTEIN
 FT Rv0278c.
 FT
 FT CONFLICT 40 40 M -> I (IN REF. 2).
 FT CONFLICT 158 163 MISSING (IN REF. 2).
 FT CONFLICT 807 807 R -> G (IN REF. 2).
 SQ SEQUENCE 957 AA; 81905 MW; 71EBAD417FBA47C CRC64;

Query Match 14.0%; Score 94.5; DB 1; Length 957;
 Best Local Similarity 32.3%; Pred. No. 2.2; Mismatches 48; Indels 23; Gaps 8;
 Matches 42; Conservative 17; Gaps 8;

QY 13 GGGFAIPIGOMAIMAIGQINSG-----GGSPYHIGPTAFGLGV-VDNNGNGARVGRVYG 66
 DB 423 GGGFAGGIGAGGTGPTGLIENGNGGNGASAVTGGNGIGGTGLIENGNGNG--SGGIG 480
 QY 67 SAPAASLIGSTGDTVAVNG--APINSA---TAMDALNGHNGDVISTVWTKSGGTR 121
 DB 481 AKRAGVGVS--GLGLIGLDFGNAPASTSPLHTLQGVNL-----VYNEFPQTLTGRPIL 532
 QY 122 GNVTLAEGPP 131
 DB 533 GNG--ANGTP 540

RESULT 13
 HQA HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable periplasmic serine protease do/hoa-like precursor
 DE (EC 3.4.21.-).
 GN H11259.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_Taxid=727;
 RN 11;
 RP SOURCE FROM N.A.
 RC STRAIN-RD / KW20 / ATCC 51907;
 RX MEDLINE=93550630; PubMed=7342800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Branton R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.,
 RA "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.,"
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA
 CC (PROTEASE DO) AND HQA.
 CC -1- SIMILARITY: CONTRAINS 2 PDZ/DHR DOMAINS.
 CC DEEP/DEGQ/DESS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U32805; AAC22906.1; -
 DR MEROPS: S01.274; -
 DR TIGR: H11259; -
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsln.
 DR Pfam: PR00595; PDZ; 2.
 DR Pfam: PR00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 KM Hydrolyase; Serine protease; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 1 466 PROBABLE PERIPLASMIC SERINE PROTEASE
 FT DO/HQA-LIKE.
 FT
 FT DOMAIN 270 361 PDZ 1.
 FT DOMAIN 367 458 PDZ 2.
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 466 AA; 49434 MW; ED050A00047B5851 CRC64;

Query Match 13.7%; Score 92.5; DB 1; Length 466;
 Best Local Similarity 27.4%; Pred. No. 1.6;
 Matches 32; Conservative 16; Mismatches 40; Indels 29; Gaps 3;

QY 13 GGGFAIPIGOMAIMA-----GQIRSGGSPYHIGPTAFGLGVVDNNGN----- 57
 DB 253 GIAFAIPNSQASNLVQOILFFGGVRRG-----LLGKGGELNDLAKAFNVSA 300
 QY 58 --GARVGRVVGSAAPASLIGSTGDTVAVNGAPINSATAMADLNGHHPDVISVTM 112
 DB 301 QGGAFVSEVLPKSAERAKGLAGDITIAMNGOKISSPAETRAKIATTGAKEISLTY 357

RESULT 14
 DEGB_ARATH STANDARD; PRT; 448 AA.
 ID DEGB_ARATH
 AC Q9L010;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protease do-like 8, chloroplast precursor (EC 3.4.21.-).
 DE DEGP8 OR AT5G39830 OR K13H13.1.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[12]
SEQUENCE OF 91-113 AND 345-361.
RC STRAIN=CV, COLUMBIA;
RA Kieselbach T., Bystedt M., Schroeder W.P.;
RL Submitted (OCT-2000) to the SWISS-PROT data bank.
CC -!- FUNCTION: PROBABLE SERINE PROTEASE.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID LUMEN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AB024023; BAA98101.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; FALSE_NEG.
KW Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
FT TRANSIT 1 ?
FT CHAIN 1 90
FT DOMAIN 91 448
FT DOMAIN 152 333
FT DOMAIN 336 433
FT ACT_SITE 171 171
FT ACT_SITE 214 214
FT ACT_SITE 292 292
FT ACT_SITE 448 AA; 47492 MW; A986FC1387670AF6 CRC64;
SEQUENCE

Query Match 13.6%; Score 92; DB 1; Length 448;
Best Local Similarity 29.5%; Pred. No. 1.6; Mismatches 55; Indels 18; Gaps 4;
Matches 36; Conservative 13; Mismatches 55; Indels 18; Gaps 4;
QY 10 SGGGFAIPIGQAMALAGQIRSGGSPVHIGPTAFGLGVVDN---NGGARVORVVG 66
Db 316 TSAGVGFAIPSTVTKIYQLIQ--FSKVLKRGINIELAPDPYANOLNVTNGALVLYQVPG 373
QY 67 SAPASALG-----ISTGVDITAVDGAIPNSATMADALNGHHPGDVISVTWQTK 115
Db 374 KSLAKKAGIHPTSRGFAIGVILDIIVAVDVKPKKAKELMKILDEYSVD--KVLTKIK 431
QY 116 SG 117
Db 432 RG 433

RESULT 15
DEGP_BUCAI STANDARD; PRT; 478 AA.
AC P57323;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).

DEGP OR BU228.
GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
[1]
SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AP001118; BAB12943.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolyase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 478
FT DOMAIN 116 254
FT DOMAIN 281 372
FT DOMAIN 387 469
FT ACT_SITE 133 133
FT ACT_SITE 163 163
FT ACT_SITE 238 238
FT ACT_SITE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;
SEQUENCE

Query Match 13.6%; Score 92; DB 1; Length 478;
Best Local Similarity 25.6%; Pred. NO. 1.7; Mismatches 50; Indels 20; Gaps 4;
Matches 32; Conservative 23; Mismatches 50; Indels 20; Gaps 4;
QY 13 GGGFAIPIGQAMALAGQIRSGGSPVHIGPTAFGLGVVD-----NNGGARVORV 64
Db 265 GIGFAIPICNMVKNLQAQVVG--QVARGELGIMGMELNSDLAQMINKSGKAFVSRV 321
QY 65 VGSAPASALGISTGVDITAVDGAIPNSATMADALNGHHPGDVISVTWQTKSGGRTTG-- 122
Db 322 LPNSAFEEAGIKADIIISLNRKPISSFSRLAETIGS-----LPVATKMEIGVPRGRI 375
QY 123 -NVTL 126
Db 376 KNITV 380

Search completed: August 6, 2002, 08:17:54
Job time: 164 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:12:10 ; Search time 16.84 Seconds
(without alignments)
191.460 Million cell updates/sec

Title: US-09-684-215A-4

Sequence: 1 TAAADNFQLSGGCGGFAIP1.....QTKSGYRTGNVTLAEGPPA 132

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PCPUS_COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Query Length | DB ID | Description |
|------------|-------|-------------|--------------|----------------------|-------------------|
| 1 | 670 | 99.3 | 132 | 4 US-08-818-112-66 | Sequence 66, Appl |
| 2 | 670 | 99.3 | 132 | 4 US-08-818-111-67 | Sequence 67, Appl |
| 3 | 670 | 99.3 | 132 | 4 US-09-056-556-66 | Sequence 66, Appl |
| 4 | 670 | 99.3 | 355 | 4 US-08-818-112-79 | Sequence 79, Appl |
| 5 | 670 | 99.3 | 355 | 4 US-08-818-111-80 | Sequence 80, Appl |
| 6 | 670 | 99.3 | 355 | 4 US-09-056-556-79 | Sequence 79, Appl |
| 7 | 200.5 | 29.7 | 580 | 4 US-08-818-112-75 | Sequence 75, Appl |
| 8 | 200.5 | 29.7 | 580 | 4 US-09-056-556-75 | Sequence 75, Appl |
| 9 | 200.5 | 29.7 | 580 | 4 US-08-818-111-76 | Sequence 76, Appl |
| 10 | 158.5 | 23.5 | 97 | 4 US-08-818-112-72 | Sequence 72, Appl |
| 11 | 158.5 | 23.5 | 97 | 4 US-08-818-111-73 | Sequence 73, Appl |
| 12 | 158.5 | 23.5 | 97 | 4 US-09-056-556-72 | Sequence 72, Appl |
| 13 | 148 | 21.9 | 30 | 4 US-09-439-313-484 | Sequence 132, App |
| 14 | 127.5 | 18.9 | 460 | 4 US-09-199-637A-132 | Sequence 132, App |
| 15 | 108 | 16.0 | 475 | 1 US-08-278-091-6 | Sequence 6, Appl |
| 16 | 108 | 16.0 | 475 | 1 US-08-483-859-6 | Sequence 6, Appl |
| 17 | 108 | 16.0 | 475 | 1 US-08-472-173-6 | Sequence 6, Appl |
| 18 | 108 | 16.0 | 475 | 1 US-08-487-167-6 | Sequence 6, Appl |
| 19 | 108 | 16.0 | 475 | 2 US-08-482-816-6 | Sequence 6, Appl |
| 20 | 108 | 16.0 | 475 | 2 US-08-296-149-6 | Sequence 6, Appl |
| 21 | 108 | 16.0 | 475 | 2 US-08-801-499-6 | Sequence 6, Appl |
| 22 | 108 | 16.0 | 475 | 2 US-08-615-271-6 | Sequence 6, Appl |
| 23 | 108 | 16.0 | 475 | 3 US-09-074-660-6 | Sequence 6, Appl |
| 24 | 108 | 16.0 | 475 | 3 US-09-074-659-6 | Sequence 6, Appl |
| 25 | 108 | 16.0 | 475 | 3 US-09-106-468-6 | Sequence 6, Appl |
| 26 | 108 | 16.0 | 475 | 4 US-09-106-468A-6 | Sequence 6, Appl |
| 27 | 108 | 16.0 | 475 | 4 US-09-106-467-6 | Sequence 6, Appl |

| | | | | | |
|----|-------|------|-----|---------------------|-------------------|
| 28 | 105 | 15.6 | 475 | 1 US-08-350-741-2 | Sequence 2, Appl |
| 29 | 105 | 15.6 | 475 | 2 US-08-463-875A-2 | Sequence 2, Appl |
| 30 | 102.5 | 15.2 | 472 | 1 US-08-278-091-5 | Sequence 5, Appl |
| 31 | 102.5 | 15.2 | 472 | 1 US-08-483-859-5 | Sequence 5, Appl |
| 32 | 102.5 | 15.2 | 472 | 1 US-08-472-173-5 | Sequence 5, Appl |
| 33 | 102.5 | 15.2 | 472 | 2 US-08-487-167-5 | Sequence 5, Appl |
| 34 | 102.5 | 15.2 | 472 | 2 US-08-482-816-5 | Sequence 5, Appl |
| 35 | 102.5 | 15.2 | 472 | 2 US-08-296-149-5 | Sequence 5, Appl |
| 36 | 102.5 | 15.2 | 472 | 2 US-08-801-499-5 | Sequence 5, Appl |
| 37 | 102.5 | 15.2 | 472 | 2 US-08-615-271-5 | Sequence 5, Appl |
| 38 | 102.5 | 15.2 | 472 | 3 US-09-074-660-5 | Sequence 5, Appl |
| 39 | 102.5 | 15.2 | 472 | 3 US-09-074-659-5 | Sequence 5, Appl |
| 40 | 102.5 | 15.2 | 472 | 3 US-09-106-468-5 | Sequence 5, Appl |
| 41 | 102.5 | 15.2 | 472 | 4 US-09-106-467-5 | Sequence 5, Appl |
| 42 | 102.5 | 15.2 | 472 | 4 US-08-923-454A-14 | Sequence 14, Appl |
| 43 | 102.5 | 15.2 | 491 | 3 US-08-350-741-3 | Sequence 3, Appl |
| 44 | 102.5 | 15.2 | 492 | 1 US-08-463-875A-3 | Sequence 3, Appl |
| 45 | 102.5 | 15.2 | 492 | 2 US-08-463-875A-3 | Sequence 3, Appl |

ALIGNMENTS

RESULT 1
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedicak, Thomas S.
; APPLICANT: Twadzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; TELEPHONE: (206) 622-4900
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 662-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-66

Query Match 99.3%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 2.4e-62;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-79

Query Match 99.3%; Score 670; DB 4; Length 355;
Best Local Similarity 99.2%; Pred. No. 8.9e-62;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFQLSGGGGAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 60
|||||
DB 224 TAASDNFQLSGGGGAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 283

QY 61 VORVGSAPASISIGTGDVITAVDGAIPINSATAMADALNGHHPGDVISTWOTKSGCTR 120
|||||
DB 284 VORVGSAPASISIGTGDVITAVDGAIPINSATAMADALNGHHPGDVISTWOTKSGCTR 343

QY 121 TGNVTLAEGPPA 132
|||||
DB 344 TGNVTLAEGPPA 355

RESULT 5
US-08-818-111-80
Sequence 80, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-80

Query Match 99.3%; Score 670; DB 4; Length 355;
Best Local Similarity 99.2%; Pred. No. 8.9e-62;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFQLSGGGGAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 60
|||||
DB 224 TAASDNFQLSGGGGAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 283

QY 61 VORVGSAPASISIGTGDVITAVDGAIPINSATAMADALNGHHPGDVISTWOTKSGCTR 120
|||||
DB 284 VORVGSAPASISIGTGDVITAVDGAIPINSATAMADALNGHHPGDVISTWOTKSGCTR 343

QY 121 TGNVTLAEGPPA 132
|||||
DB 344 TGNVTLAEGPPA 355

RESULT 6
US-09-056-556-79
Sequence 79, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392

```
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-036-536-79

Query Match
Best Local Similarity 99.3%; Score 670; DB 4; Length 355;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVNNNGNGAR 60
DB 224 TAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVNNNGNGAR 283
QY 61 VORVGSAPASISGISTGDVITAVDGAIPINSATMADALNGHHPGDVTSVTWQTKSGTR 120
DB 284 VORVGSAPASISGISTGDVITAVDGAIPINSATMADALNGHHPGDVTSVTWQTKSGTR 343
QY 121 TGNVTLAEGPPA 132
DB 344 TGNVTLAEGPPA 355

RESULT 7
US-08-818-112-75
Sequence 75, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
```

```
TOPOLOGY: linear
US-08-818-112-75

Query Match
Best Local Similarity 29.7%; Score 200.5; DB 4; Length 580;
Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TAASDNFQLSQG--GGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVV--DNNGN 57
DB 453 TLGADSDAOSGSLGLGFALPVDARKIADDELISTGKA-----SHASLGVQVYNDKDP 506
QY 58 GARVGRVGSAPASISGISTGDVITAVDGAIPINSATMADALNGHHPGDVTSVTWQTKSG 117
DB 507 GAKIVEVAGGAANAANQVPRGVYTVKDDRPINSADALVAAYVSKAPGATVALTFDDPSG 566
QY 118 GTRTGNVTLAE 128
DB 567 GSRTVQVTLGK 577

RESULT 8
US-08-818-111-76
Sequence 76, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-76

Query Match
Best Local Similarity 29.7%; Score 200.5; DB 4; Length 580;
Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TAASDNFQLSQG--GGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVV--DNNGN 57
```

D_b 453 TLGADSDAQSGSTIGLGFALPVDQAKRIADELLISTGCA-----SHAHSLSGVYTNKDDTP 506

Q_y 58 GARQRVRVGCNAPASLCISITGDYTANTDAGPINSATRTMADALNGHEHGVISTWTQTSCG 117
||| ||| : | : | ||| ||| : : : |||

D_b 507 GAKIYEVEVAGGAANAGVPRGVVTKVDRPINSADALVAARSKAPGATVALFPDPDSG 566
||||| ||| :

Q_y 118 GTRTGNVTLAEL 128
| : ||| ||| :

D_b 567 GSRTGVQVTLGR 577

RESULT 9
US-09-056-556-75

```

: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Dillon, Davin C.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
: NUMBER OF SEQUENCES: 241
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/056,556
: FILING DATE: 07-APR-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: MakI, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.457
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 75:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 580 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IS-09-056-556-75

```

| | | | | |
|-----------------------|-----------------|-------------------|----------|------------|
| Query Match | 29.7% | Score 200.5 | DB 4 | Length 580 |
| Best Local Similarity | 38.9% | Pred. No. 7.7e+13 | | |
| Matches 51 | Conservative 21 | Mismatches 50 | | |
| | | | Indels 9 | Gaps 3 |

RESULT 10
US-08-818-112-72
; Sequence 72, Application US/08818112

Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

| | | | | |
|-----------------------|------------------|--------------------|-----------|------------|
| Query Match | 23.5%; | Score 158.5; | DB 4; | Length 97; |
| Best Local Similarity | 38.7%; | Pred. No. 1.7e-09; | | |
| Matches 36; | Conservative 16; | Mismatches 38; | Indels 3; | Gaps 1; |

```

QY      96 ADALNGHHPGDVISVTWQTKSGGTRGNVTLAE 128
      11  ::::: 11111 111  :
DB      62 VAAVRSKAPGATVALTFQDPSSGSRVQVYTLGK 94

```

```

RESULT 11
US-08-818-111-73
: Sequence 73, Application US/08818111
: Patent No. 6338852
:
: GENERAL INFORMATION:
:
: APPLICANT: Reed, Steven G.
:
: APPLICANT: Skelky, Yasir A.W.
:
: APPLICANT: Dillon, David C.
:
: APPLICANT: Campos-Neto, Antonia
:
: APPLICANT: Houghton, Raymond
:
: APPLICANT: Vedvick, Thomas S.
:
: APPLICANT: Twardzik, Daniel R.
:
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
: TUBERCULOSIS
:
: NUMBER OF SEQUENCES: 148
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: SEED and BERRY LLP

```

STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MakI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IS-08-818-111-73

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:14:20 ; Search time 26.07 Seconds
(without alignments)
875.924 Million cell updates/sec

Title: US-09-684-215A-4

Perfect score: 675
Sequence: 1 TAAADNFQSLQSGGQGFAPL.....QTKSGSTRGNVTLACPPA 132

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SPTREMBL_19:**
Listing first 45 summaries

1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mhc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_virus:**
16: sp_bacterioph:**
17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|---------------------|
| 1 | 675 | 100.0 | 355 | 16 0071175 | 0071175 mycobacteri |
| 2 | 474.5 | 70.3 | 361 | 2 050320 | 050320 mycobacteri |
| 3 | 459.5 | 68.1 | 354 | 16 09C9C9 | 09C9C9 mycobacteri |
| 4 | 199.5 | 28.6 | 464 | 16 053896 | 053896 mycobacteri |
| 5 | 188.5 | 27.9 | 382 | 16 09CDE7 | 09CDE7 mycobacteri |
| 6 | 188.5 | 27.9 | 452 | 2 0925G6 | 0925G6 mycobacteri |
| 7 | 161.5 | 23.9 | 542 | 2 09FBK9 | 09FBK9 streptomyc |
| 8 | 154 | 22.8 | 519 | 2 09J7J0 | 09J7J0 streptomyc |
| 9 | 139 | 20.6 | 394 | 16 P72780 | P72780 streptomyc |
| 10 | 138.5 | 20.5 | 362 | 2 09S2R5 | 09S2R5 streptomyc |
| 11 | 131 | 19.4 | 441 | 16 09RTK4 | 09RTK4 streptomyc |
| 12 | 128.5 | 19.0 | 514 | 16 09PBA3 | 09PBA3 xyllella fas |
| 13 | 127.5 | 18.9 | 371 | 2 031388 | 031388 bradyrhizob |
| 14 | 127.5 | 18.9 | 474 | 2 09ALST | 09ALST pseudomonas |
| 15 | 122.5 | 18.1 | 474 | 16 057155 | 057155 pseudomonas |
| 16 | 121 | 17.9 | 513 | 16 098CS8 | 098CS8 rhizobium 1 |

| | | | | | |
|----|-------|------|-----|-----------|---------------------|
| 17 | 120.5 | 17.9 | 465 | 16 0920E6 | 0920E6 rhizobium m |
| 18 | 119.5 | 17.7 | 203 | 2 09X840 | 09X840 streptomyc |
| 19 | 118 | 17.5 | 516 | 16 0985F9 | 0985F9 rhizobium 1 |
| 20 | 117.5 | 17.4 | 530 | 16 09A4S2 | 09A4S2 caulobacter |
| 21 | 117 | 17.3 | 473 | 2 044476 | 044476 azotobacter |
| 22 | 113 | 16.7 | 393 | 16 097N37 | 097N37 streptococ |
| 23 | 113 | 16.7 | 397 | 2 006670 | 006670 streptococ |
| 24 | 112 | 16.6 | 500 | 2 044652 | 044652 bruceella ab |
| 25 | 111 | 16.4 | 481 | 2 09AQD1 | 09AQD1 pseudomonas |
| 26 | 111 | 16.4 | 499 | 16 09JVT1 | 09JVT1 neisseria m |
| 27 | 108 | 16.0 | 348 | 16 097LU1 | 097LU1 clostridium |
| 28 | 108 | 16.0 | 472 | 16 09A8R9 | 09A8R9 caulobacter |
| 29 | 107.5 | 15.9 | 452 | 16 P73354 | P73354 streptocyst |
| 30 | 104.5 | 15.5 | 481 | 2 09L8K0 | 09L8K0 shigella so |
| 31 | 104 | 15.4 | 397 | 2 093F87 | 093F87 streptococ |
| 32 | 103.5 | 15.3 | 407 | 16 095XG9 | 095XG9 streptococ |
| 33 | 102.5 | 15.2 | 428 | 16 098N31 | 098N31 rhizobium 1 |
| 34 | 102 | 15.1 | 500 | 2 09KJN6 | 09KJN6 myxococcus |
| 35 | 100.5 | 14.9 | 453 | 16 067436 | 067436 aquifex ao |
| 36 | 100.5 | 14.9 | 459 | 16 09WZ41 | 09WZ41 thermotoga |
| 37 | 100 | 14.8 | 398 | 16 083752 | 083752 treponema p |
| 38 | 100 | 14.8 | 652 | 16 092G62 | 092G62 rickettsia |
| 39 | 99.5 | 14.7 | 224 | 16 035021 | 035021 bacillus su |
| 40 | 99.5 | 14.7 | 458 | 2 09R8T1 | 09R8T1 bacillus su |
| 41 | 98.5 | 14.6 | 481 | 16 09PGL3 | 09PGL3 xyllella fas |
| 42 | 98 | 14.5 | 476 | 16 09ZM18 | 09ZM18 helicobacte |
| 43 | 97.5 | 14.4 | 453 | 2 09FDP1 | 09FDP1 aeromonas h |
| 44 | 95.5 | 14.1 | 290 | 5 021294 | 021294 caenorhabdi |
| 45 | 95.5 | 14.1 | 404 | 16 033351 | 033351 mycobacteri |

ALIGNMENTS

| RESULT | ID | PRELIMINARY: | PRT: | 355 AA. |
|--------|--|--------------|------|---------|
| 007175 | 007175 | | | |
| AC | 007175 | | | |
| DT | 01-JUL-1997 (TREMBlrel. 04, Created) | | | |
| DT | 01-JUL-1997 (TREMBlrel. 04, Last sequence update) | | | |
| DT | 01-DEC-2001 (TREMBlrel. 19, Last annotation update) | | | |
| DE | HYPOTHETICAL 34.9 KDA PROTEIN. | | | |
| GN | PEPA OR RV0125 OR MTC1418B.07. | | | |
| OS | Mycobacterium tuberculosis. | | | |
| OC | Bacteria: Firmicutes: Actinobacteria: Actinobacteridae: | | | |
| OC | Actinomycetales: Corynebacterineae: Mycobacteriaceae: Mycobacterium. | | | |
| OX | NCBI_TaxID=1773; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=H37RV; | | | |
| RX | MEDLINE=98295987; PubMed=9634230; | | | |
| RA | Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., | | | |
| RA | Badcock K., Basham D., Brown D., Chillingworth T., Connor R., | | | |
| RA | Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., | | | |
| RA | Hornby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., | | | |
| RA | Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., | | | |
| RA | Rutten S., Seeger K., Skelton S., Squares S., Squares R., | | | |
| RA | Sutten J.E., Taylor K., Whitehead S., Barrell B.G.; | | | |
| RT | "Deciphering the biology of Mycobacterium tuberculosis from the | | | |
| RP | complete genome sequence." | | | |
| RL | Nature 393:537-544(1998). | | | |
| CC | EMBL: Z96071; CAB09453.1; -. | | | |
| DR | MEROPS: S01.0PC; -. | | | |
| DR | TubercuList: RV0125; -. | | | |
| DR | InterPro: IPR001478; PDZ. | | | |
| DR | InterPro: IPR001254; Trypsin. | | | |
| DR | Pfam: PF00595; PDZ; 1. | | | |
| DR | Pfam: PF00088; trypsin; 1. | | | |
| DR | SMART: SM00228; PDZ; 1. | | | |
| DR | PROSITE: PS50106; PDZ; 1. | | | |
| DR | PROSITE: PS50240; TRYPSIN_DOM; 1. | | | |

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN.1.
 KM Complete proteome: Hydrolase; Hypothetical protein; Serine protease.
 SQ SEQUENCE 355 AA; 34926 MW; 16CB9E21A97BF192 CRC64;

Query Match 100.0%; Score 675; DB 16; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1.5e-41;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPGOMALTAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 DB 224 TAASDNFOLSGGGGFAIPGOMALTAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
 QY 61 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVSVTWQTKSGGTR 120
 DB 284 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVSVTWQTKSGGTR 343
 QY 121 TGNVTLAEGPPA 132
 DB 344 TGNVTLAEGPPA 355

RESULT 2
 Q50320 PRELIMINARY; PRT; 361 AA.
 AC 050320;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 34KDA PROTEIN PRECURSOR.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JDS8/107;
 RX MEDLINE=95003449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterization of a putative serine protease
 expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
 RL Microbiology 140:1977-1982(1994).
 DR EMBL: 223092; CAA80638.1; -;
 DR MEROPS: S01.UPC; -;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ.1.
 DR Pfam: PF00089; Trypsin.1.
 DR SMART: SM00228; PDZ.1.
 DR PROSITE: PS50106; PDZ.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR Hydrolase; Serine protease; Signal.
 KW SIGNAL.1
 FT SIGNAL.1
 SQ SEQUENCE 361 AA; 35709 MW; 30EFB78FDF6F3C411 CRC64;

Query Match 70.3%; Score 474.5; DB 2; Length 361;
 Best Local Similarity 69.7%; Pred. No. 4.9e-27;
 Matches 92; Conservative 17; Mismatches 22; Indels 1; Gaps 1;

QY 1 TAASDNFOLSGGGGFAIPGOMALTAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 DB 231 TAATDSTKMS-GGQGFALPIGRMAVAANQIRSGAGSNTVHIGPTAFGLGVVDNNGNGAR 289
 QY 61 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVSVTWQTKSGGTR 120
 DB 290 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVSVTWQTKSGGTR 349
 QY 121 TGNVTLAEGPPA 132
 DB 350 TGNVTLAEGPPA 361

RESULT 3
 Q9CCY9 PRELIMINARY; PRT; 354 AA.
 ID Q9CCY9;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PROBABLE SECRETED SERINE PROTEASE.
 GN M2659.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN.
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajadream M., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AL583926; CAC32191.1; -;
 DR MEROPS: S01.UPC; -;
 DR Leprosin; M2659; -;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR000126; Ser protease_v8.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ.1.
 DR Pfam: PF00089; Trypsin.1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR PRINTS: PR00839; V8PROTEASE.
 DR SMART: SM00228; PDZ.1.
 DR PROSITE: PS50106; PDZ.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR Hydrolase; Serine protease; Signal.
 KW SIGNAL.1
 SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9E4A4 CRC64;

Query Match 68.1%; Score 459.5; DB 16; Length 354;
 Best Local Similarity 66.7%; Pred. No. 5.9e-26;
 Matches 88; Conservative 17; Mismatches 26; Indels 1; Gaps 1;

QY 1 TAASDNFOLSGGGGFAIPGOMALTAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 DB 224 TAATDNTKM-LGGGFAIPGOMAEVGAIRSGAGSNTVHIGPTAFGLGVVDNNGNGAR 282
 QY 61 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVSVTWQTKSGGTR 120
 DB 283 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVSVTWQTKSGGTR 342
 QY 121 TGNVTLAEGPPA 132
 DB 343 TGNVTLAEGPPA 354
 RESULT 4
 O53896 PRELIMINARY; PRT; 464 AA.
 ID O53896;
 AC O53896;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE SERINE PROTEASE.
 GN RV0983 OR MTV044.11.

OC Mycobacteria tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxId:1773;
 RN [1]
 RP
 RC
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RX Cole S.T., Brosch R., Parthall J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Mekala F.,
 RA Badcock R., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Kiroh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rulten S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulton J.E., Taylor K., Whitehead S., Barrett L.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RT Nature 393:537-544(1998).
 RL
 RL -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR ENBL: AL021999; CAA17562.1; -.
 DR MEROPS: S01.0PC;-;
 DR TubercuList; RV0983;-;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ: 1.
 DR Pfam: PF00089; trypsin; 1.
 DR SMART: SMO0228; PDZ: 1.
 DR PROSITE: PSS50106; PDZ: 1.
 DR PROSITE: PSS50240; TRYPSIN_DOM: 1.
 DR Complete proteome; Hydrolase; Protease; Serine protease.
 QO SEQUENCE 464 AA; 46452 MW; A693BFC531EEC8F CMC64;

| | | | | | |
|-----------------------|--|--|--------------------|------------|-------------|
| Query Match | | 29.6% | Score 199.5; | DB 16; | Length 466; |
| Best Local Similarity | | 38.9% | Pred. No. 5.4e-07; | | |
| Matches | 51; | Conservative | 21; | Mismatches | 50; |
| | | | | Indels | 9; |
| | | | | Gaps | 3. |
| OY | 1 | TAASNPFLDSOG--GCGPAIRPGAMATAGCIRSGGSPTYHIGTPAFLGLGV--DNNGN | 57 | | |
| | | : : | : : : | : : | |
| Dd | 337 | TLCGADSDAOGSGISLGLAIRPYDAKRRLADELISTGKA-----SHASLGVOYTNDKDTL | 390 | | |
| OY | 58 | GARYORVGSAPASLSLSTGDVTITAVDGAPINSATAMADALNGHHPPGDVISVTWOTKSG | 117 | | |
| | | : : | : : | : : | |
| Dd | 391 | GAKIYEVAAGAANAAGVPKGVVYTKVDPRINSADLVAAVRSAKAPATVALTFQDPGSG | 450 | | |
| OY | 118 | GTRIGNVTPLAE | 128 | | |
| | | : : | | | |
| Dd | 451 | GSRTQVYTLGK | 461 | | |
| RESULT | 5 | | | | |
| QCDC67 | | | | | |
| ID | QCDC67 | PRELIMINARY; | PRT; | 382 AA. | |
| AC | QCDC67; | | | | |
| DT | 01-JUN-2001 (TREMBLrel. 17; Created) | | | | |
| DR | 01-JUN-2001 (TREMBLrel. 17; Last sequence update) | | | | |
| DE | 01-DEC-2001 (TREMBLrel. 19; Last annotation update) | | | | |
| DE | POSSIBLE SECRETED SERINE PROTEASE. | | | | |
| GN | ML0176. | | | | |
| OS | Mycobacterium leprae. | | | | |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | | | | |
| CC | Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium. | | | | |
| OX | NCBI_TaxID=1769; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN-TN; | | | | |
| RX | MEDLINE=21128732; PubMed=11234002; | | | | |
| RA | Cole S.T., Elgmeier K., Parkhill J., James K.D., Thomson N.R., | | | | |
| RA | Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D., | | | | |
| RA | Mungall K., Basham D., Brown D., Chillingworth T., Connor R., | | | | |
| RA | Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N., | | | | |
| RA | Holtroyd S., Hornby T., Jagels K., Lacroix C., Maclean J., Moule S., | | | | |

RA Murphly L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Siml S., Simmonds M., Skellton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrett B.G.: "Massive gene decay in the leprosy bacillus.",
RT Nature 409:1007-1011(2001).
RL -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC EMBL; AL583917; CAC29684.1; -.
DR MEROPS; S01.0PC; -.
DR Lepidoma; ML0176; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PR00595; PDZ; 1.
DR Pfam; PF00069; trypsin; 1.
DR PRINTS; PR00834; PROTEASE2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR Complete proteome; Hydrolyase; Protease; Serine protease.
KW SEQUENCE 382 AA; 37084 MW; 3DD8DDDAE32A80D CRC64;

```

Query Match          27.9%; Score 188.5; DB 16; Length 382;
Best Local Similarity 39.8%; Pred. No.2.7e-06;
Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

Oy 2 AASDFQFQSSGGGCGFAIPGQAMAIAGQIRSGSGSPYHIGPTAFGLGVVDNNGN-GAR 60
   | : | : | | | | : | | : | : | : | : | : | : | : | : | : | : |
Db 258 ADSGAGQASIGLGRFATVPDQAKRRRLADELSIG--KATN----ASLGQVAVDKKTPGAK 311
Oy 61 VQRVVGSAPPAASLGISTGCVITAVDGAIPNATYAMADLNGHHPCDVISVTWQYSGGTR 120
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 312 VMDVYAGGAAANAAPVKGVLTFRKVDRLTSSADALVAAVRSKAPGDKVSLTYQDQSGSSR 371
Oy 121 TGNVTLAE 128
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 372 TVQVTLGK 379

RESULT 6
ID 0925G6 PRELIMINARY; PRT; 452 AA.
AC Q925G6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE SERINE PROTEASE.
GN MLCB373.28.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1169;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris D., Taylor K.;
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eigmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
RT of Mycobacterium leprae.";
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL; AL035500; CAB36690.1; -.
DR MEROPS; S01.UPC; -.
DR InterPro; IPR001478; PDX.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PD2; 1.

```



```

P72780
ID P72780 PRELIMINARY; PRT: 394 AA.
AC P72780;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE PROTEASE_HHOA.
GN HHOA OR SL11679.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: D90900; BAA16795.1;
DR MEROPS: S01.274;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ. 1.
DR Pfam: PF00089; trypsin. 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ. 1.
DR PROSITE: PS50106; PDZ. 1.
DR PROSITE: PS50240; TRYPSIN_DOM. 1.
DR Complete proteome: Hydrolase; Protease; Serine protease.
KW SEQUENCE 394 AA; 41336 MW; AD59D94811B8F57B CRC64;
SQ

Query Match 20.6%; Score 139; DB 16; Length 394;
Best Local Similarity 34.6%; Pred. No. 0.011;
Matches 45; Conservative 10; Mismatches 43; Indels 32; Gaps 5;

OY 13 GGGFAIPGQAMAIAGQIRSGGSPYVHIGPTAFGLGIV---VDNNGN-----G 58
DB 261 GIGFAIPIDAKAIQNTLAAGTVPHPYIG-VQMMNITVDQAQNNRNPNSEFTIPEDVG 319
OY 59 ARVQRVVGSAPASLIGSTGDIVITAVDGAIPINSATAM-----ADALNGHHP 104
DB 320 ILVMRYLPGTPAERAGIRRGDIVIAYVDGTPISDGA RLQRIVEQAGLNKALKDLRLGDRR 379
OY 105 GDVISTVWQT 114
DB 380 ---LSLTVQT 386

RESULT 10
O9S2R5 PRELIMINARY; PRT: 362 AA.
AC O9S2R5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE SECRETED PROTEASE.
GN SC5F7.30.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K., Harris D.;

```

```

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleiser H.M., Denapite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL096872; CAB51255.1;
DR MEROPS: S01.UFC;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ. 1.
DR Pfam: PF00089; trypsin. 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ. 1.
DR PROSITE: PS50106; PDZ. 1.
DR PROSITE: PS50240; TRYPSIN_DOM. 1.
DR Hydrolase; Protease; Serine protease.
KW SEQUENCE 362 AA; 36381 MW; 71CC41F943A6D03B CRC64;
SQ

Query Match 20.5%; Score 138.5; DB 2; Length 362;
Best Local Similarity 35.5%; Pred. No. 0.011;
Matches 43; Conservative 14; Mismatches 55; Indels 9; Gaps 4;

OY 13 GGGFAIPGQAMAIAGQIRSGG---GSPYVHIGPTAFGLGVNNGN--GARVQRVVGS 67
DB 245 GIGFAIPASVTVYAGQIVADGKVTDSGRALGITA---RTVVDSTYRPAANAAYVESDG 301
OY 68 APASLIGSTGDIVITAVDGAIPINSATAMADALNGHHPGDIVSTVWQTKSGGTRGNVTLA 127
DB 302 GAADDAGLRGPDVLVKLGDYDITITLSLSEALASMRGCDPRKVTY-TRDGKEHRAEVTIG 360
OY 128 E 128
DB 361 E 361

RESULT 11
O9RTK4 PRELIMINARY; PRT: 441 AA.
AC O9RTK4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PERIPLASMIC SERINE PROTEASE DO, PUTATIVE.
GN DR1756.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).

```

CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AE002017; AAF1312.1; -
 DR MEROPS: S01.274; -
 DR TIGR: DRI756; -
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR Complete proteome: Hydrolase; Serine protease.
 KW SEQUENCE 441 AA; 45145 MW; D9E75AD081E0190 CRC64;
 SQ

Query Match 19.4%; Score 131; DB 16; Length 441;
 Best Local Similarity 29.0%; Pred. No. 0.047;
 Matches 47; Conservative 24; Mismatches 43; Indels 48; Gaps 6;

OY 6 NFOLOGGCG-----GFAIPIGQAMAIAGIIRSGGSPYTHIGPTAFLGLGVYDNN----- 55
 DB 277 NFOILTGAGAGSAGVGFAIPINTVKRLPDLQAGKG-----GVSPSPSLGVFSDLSL 330
 OY 56 -----GNGARVQHVVGSPAPASLG-----IST-GDVITAV 84
 DB 331 PQQLKAAAGLPGSGALLQGVYPPSPAPAAAGLRGNGNKGKLSLPSAQGTSSTSDGLITAV 390
 OY 85 DGAPINSATAMADALNGHHHPGVISTWTQKSGCTRGNTL 126
 DB 391 NQOPLDAGSLQEAVALTGEQPLRLTVR-RGKTRREVETL 431

RESULT 12
 O9PBA3 PRELIMINARY; PRT; 514 AA.
 ID O9PBA3;
 AC O9PBA3;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PERPLASMIT PROTEASE.
 GN XF2241.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OC NCBI_TaxID=2371;
 RN NCBI_TaxID=2371;
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Aryda J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,
 RA Bueno M.R.P., Colombo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhami A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zeldanis J., Zetbal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."
 RL Nature 406:151-159(2000).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AE004037; AAF85040.1; -
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR000126; Ser-proteas-V8.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR PRINTS: PR00839; VBPROTEASE.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR Complete proteome: Hydrolase; Serine protease.
 KW SEQUENCE 514 AA; 54140 MW; 707C23FD3C82BE4C CRC64;
 SQ

Query Match 19.0%; Score 128.5; DB 16; Length 514;
 Best Local Similarity 38.7%; Pred. No. 0.084;
 Matches 43; Conservative 11; Mismatches 46; Indels 11; Gaps 4;

OY 4 SDNFOLOGGCG--GOGFAIPIGQAMAIAGIIRSGC-----GSPYTHIGPTAFL--GLGVYDNN 54
 DB 262 SQIFASAGGYMCISFAIRININAINAEQIRKTKGYORSLMGVEIGIDAKKAGGLDLPSS 321
 OY 55 NNGARVQHVVGSPAPASLGISTGVITAVDGAIPINSATAMADALNGHHHPG 105
 DB 322 --RGALVNNIPPHSPAPAKAGLEVGVIRSVNGKVISFSFDLPPLIGMPPG 370

RESULT 13
 O31388 PRELIMINARY; PRT; 371 AA.
 ID O31388;
 AC O31388;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE DEGP PROTEIN.
 GN DEGP.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OC NCBI_TaxID=375;
 RN NCBI_TaxID=375;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98114461; PubMed=9446679;
 RA Nardorhaus F., Weighofer W., Fischer H.M., Hennecke H.;
 RT "Identification of the Bradyrhizobium japonicum depp gene as part of
 an operon containing small heat shock protein genes."
 RL Arch. Microbiol. 168:89-97(1998).
 DB EMBL: Y13616; CAAT7338.1; -
 DB HSSP: P31016; 1BFE.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 371 AA; 39108 MW; 3CE1C77E17B86CDC CRC64;
 SQ

Query Match 18.9%; Score 127.5; DB 2; Length 371;
 Best Local Similarity 28.9%; Pred. No. 0.065;

| | | | | | | | | | |
|-----------------------|---|--|-----------------|-------------|-----|--------|-----|------|---|
| Matches | 33; | Conservative | 24; | Mismatches | 55; | Indels | 7; | Gaps | 4 |
| QY | 13 | GCGFAIRIGQAMALAGQIRSGGSPYHIGTFAFLGCV--- | VDNNGARVQRVVGAP | 69 | | | | | |
| DB | 243 | GIGFAVPRINMARWRWEDLVQYGEVRRQIG-ISTRIDVDLAKESSYGALLAETIASSP | 301 | | | | | | |
| QY | 70 | AASGISTGDTVTAVDGAIPNSATAMADALGNHPRGVISTWQTSGGTFRGNTYLAEG | 129 | | | | | | |
| DB | 302 | AEQAGLQKGDLYKAVNDGPRIRASQRLNLIGLTPVGSVELRFE-RNQAARSASVEV--G | 358 | | | | | | |
| QY | 130 | P 130 | | | | | | | |
| DB | 359 | P 359 | | | | | | | |
| RESULT | 14 | | | | | | | | |
| Q9ALSL1 | | PRELIMINARY; | PRT; | 474 | AA. | | | | |
| AC | Q9ALSL1 | | | | | | | | |
| DT | 01-JUN-2001 | (TREMBLrel. 17, Created) | | | | | | | |
| DT | 01-JUN-2001 | (TREMBLrel. 17, Last sequence update) | | | | | | | |
| DT | 01-DEC-2001 | (TREMBLrel. 19, Last annotation update) | | | | | | | |
| DE | MUCD. | | | | | | | | |
| GN | MUCD. | | | | | | | | |
| OS | Pseudomonas aeruginosa. | | | | | | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; | | | | | | | | |
| OC | Pseudomonas. | | | | | | | | |
| OX | NCBI_TaxID=287; | | | | | | | | |
| RN | [1] | | | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | | | | |
| RC | STRAIN=UCBPP-PA14; | | | | | | | | |
| RA | Yorjey P.S., Raimo L.G., Tan M., Ausubel F.M.; | | | | | | | | |
| RT | "The Roles of mucD and Alginate in the Virulence of Pseudomonas | | | | | | | | |
| RT | aeruginosa (Jan-2001) to the EMBL/Genbank/DBJ databases. | | | | | | | | |
| RL | Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases. | | | | | | | | |
| DR | EMBL: AF343973; AAK1276.1; " | | | | | | | | |
| DR | InterPro: IPR001478; PDZ. | | | | | | | | |
| DR | InterPro: IPR001940; Protease2C. | | | | | | | | |
| DR | InterPro: IPR001254; Trypsin. | | | | | | | | |
| DR | Pfam: PF00595; PDZ; 2. | | | | | | | | |
| DR | Pfam: PF00089; trypsin; 1. | | | | | | | | |
| DR | PRINTS: PR00834; PROTEASES2C. | | | | | | | | |
| DR | SMART: SM00228; PDZ; 2. | | | | | | | | |
| DR | PROSITE: PS0106; PDZ; 2. | | | | | | | | |
| DR | PROSITE: PS0240; TRYPsin_DOM; 1. | | | | | | | | |
| DR | Hydrolase; Serine protease. | | | | | | | | |
| SO | SEQUENCE 474 AA; 50349 MW; A22FD4338B859D4C CRC64; | | | | | | | | |
| Query Match | 18.9%; | Score 127.5; | DB 2; | Length 474; | | | | | |
| Best Local Similarity | 31.2%; | Pred. No. 0.09; | | | | | | | |
| Matches | 43; | Conservative | 21; | Mismatches | 51; | Indels | 23; | Gaps | 5 |
| QY | 4 | SDNFQLSOG--GCGFAIRIGQAMALAGQIRSGGSPYHIGTFAFLGCVVDNN----- | 55 | | | | | | |
| DB | 233 | SGIFRSGGFMGLSPFIPIDVALNVAADDLKKAG--KVSRG--WLGVIYIQEYNNKDLAES | 286 | | | | | | |
| QY | 56 | -----GNGARVQRVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALGNHPRGVISY | 110 | | | | | | |
| DB | 287 | FGLDKPRSCSALVAQLVEDPAAAGGLQGVYILSLNGQSI NESADLPRLHGNMKRPGDKINL | 346 | | | | | | |
| QY | 111 | ----TWQTSGGTFRGNV | 124 | | | | | | |
| DB | 347 | DVIRNGQRRKSLSMAYGNL | 364 | | | | | | |
| RESULT | 15 | | | | | | | | |
| Q57155 | | PRELIMINARY; | PRT; | 474 | AA. | | | | |
| AC | Q57155 | | | | | | | | |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Created) | | | | | | | |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Last sequence update) | | | | | | | |
| DT | 01-DEC-2001 | (TREMBLrel. 19, Last annotation update) | | | | | | | |

DE MUCD (SERINE PROTEASE MUCD).
GN MUCD OR PA0766.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=93391358; PubMed=8378309;
RA Martin D.W., Schurr M.J., Mudd M.H., Govan J.R., Holloway B.W.,
RA Deretic V.;
RT "Mechanism of conversion to mucoidy in Pseudomonas aeruginosa
RT infecting cystic fibrosis patients";
RT proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).
RL [2]
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=95286510; PubMed=7768826;
RA Yu H., Schurr M.J., Deretic V.;
RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas
RT aeruginosa AlgU: E. coli rpoE restores mucoidy and reduces sensitivity
RT to reactive oxygen intermediates in algU mutants of P. aeruginosa.";
RL J. Bacteriol. 177:3259-3268(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=96134987; PubMed=8550474;
RA Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd M.H., Yu H.,
RA Deretic V.;
RT "Two distinct loci affecting conversion to mucoidy in Pseudomonas
RT aeruginosa in cystic fibrosis encode homologs of the serine protease
RT HtrA";
RL J. Bacteriol. 178:511-523(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Britkhan F.S.U., Hutnagel W.O., Kowalik D.J., Lagrou M.,
RA Gader R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Foltner K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen";
RT Nature 406:959-964(2000).
CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: U949151; AAC43718.1; -
DR EMBL: U32853; AAC43676.1; -
DR EMBL: AE004511; AAG04155.1; -
DR MEROPS: S01.UFC; -
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; Hydrolase; Protease; Serine protease.
SO SEQUENCE 474 AA; 50321 MW; 8AB5D2A89867BEBE CRC64;

[illegible]

Search completed: August 6, 2002, 08:17:32
Job time: 192 sec